International Journal of Future Generation Communication and Networking
Vol. 13, No. 3, (2020), pp. 1352–1364

Cardiac disease related parameters analysis and disease prediction with health tips support using data mining and deep learning

Chetan Andhare[1], Musharraf Shaikh[2], Suyash Vartak[3], Saurabh Patil[4], Prasad Khawale[5], Dr. D. R. Ingale[6]

[1]*Faculty, Dept of Information Technology, Government College of Engineering Karad, India*

[2,3,4,5] *B. Tech., Dept of Information Technology, Government College of Engineering Karad, India*

[6]*Faculty, Bharati Vidyapeeth College of Engineering, Navi Mumbai, India*

Abstract

Today the main reason behind deaths in the world is the heart disease. For citizens, medical experts and practitioners, the critical task is to predict the heart disease due to unawareness of symptoms that causes the major problem to the human being. In India, awareness can be improved in the citizens and medical practitioners by providing deeper training and learning for the better human life. Despite heart function monitoring is the very essential extent of analysis and estimation that is the impact cause for the stage of heart with the other factors of heart healthiness like ECG, cholesterol, blood pressure and level of sugar. In previous research conducted on same issue, 13 features were used to predict heart disease. The results obtained in the previous work shows accuracy up to 93.33%. The research conducted by our group considering same features after changing the hyper parameters the accuracy was improved up to 95.25%. The basic objective of this is to develop a cost-effective application using deep learning technologies that will able to identify the cardiac disease severity for facilitating decision support system with health tips and suggestion. This paper explains the prediction system for monitoring heart fitness related parameters and described how to use deep learning algorithm and techniques like deep neural network to predict the heart disease in four stages.

Keywords: *Data mining, deep neural network, deep learning, heart disease, blood parameters, flask.*

1. Introduction

Health is one of the comprehensive threats for society. Huge population and lack of digital awareness in India are the biggest problems to serve the health care facilities for society. It is a basic need of every citizen to get a good healthcare support. At present the method with good accuracy and precision for the diagnosis of the heart diseases is the Angiography, which is commonly used by the physicians [7]. But it is costly and has major side effects and is difficult to consider all parameters. These problems further lead to the development of heart disease detection system. There are some methods which results human errors and are irrelevant to patient parameters like patient's medical history [9], [8]. As technology emerge, it is possible to automate medical instruments, data analysis, monitoring, prediction and providing support which will help for better diagnosis.[5] Smart phones, wearable devices and now easily available. Internet sources can make health care system more popular and available. Deep learning refers to employing techniques to process knowledge and extracts information which can be used to predict the various stages of the heart diseases, from initial stage to critical stage. The health care system collects huge amounts of health care knowledge that is for effective predictions and decision making. Discovering relations that connect variables in an exceedingly database is that the subject of knowledge mining. This work aims

in developing a prediction support in cardiopathy prediction system, exploitation data processing, modelling techniques, namely deep learning. Medical parameters required to check the probability of heart diseases where as it may be normal or critical heart condition requires citizen age, male or female, varying heart rate, ECG parameters, blood pressure range and blood sugar. Exploitation data processing information, techniques and clustering, association analysis can build additionally a prediction report for future diseases of a patient. Government or several authorities might use this analysis and may take steps to cure the disease. Different prediction models based on feature pre-processing and deep learning have been developed and proposed to improve the accuracy in diagnosis of patient. These systems have been developed to improve the accuracy and to predict whether a particular patient has a heart disease or not. Study of different diagnosis system motivates us to develop and to redesign existing and new systems. This type of system overcomes both overfitting and underfitting. In this paper a DNN based heart disease prediction model is proposed. It is an automated medical diagnosis model. This model is used for the early prediction of the heart disease. The performance of the model is evaluated on the Cleveland dataset which offers 95.25% accurate results.

In the following FIGURE 1, the pre-processing block represents the normalization process which is carried out on the featured vectors. The dataset is then divided into the train data and the test data. The 13 features are selected for the processing which are used while training as well as for testing of the model.

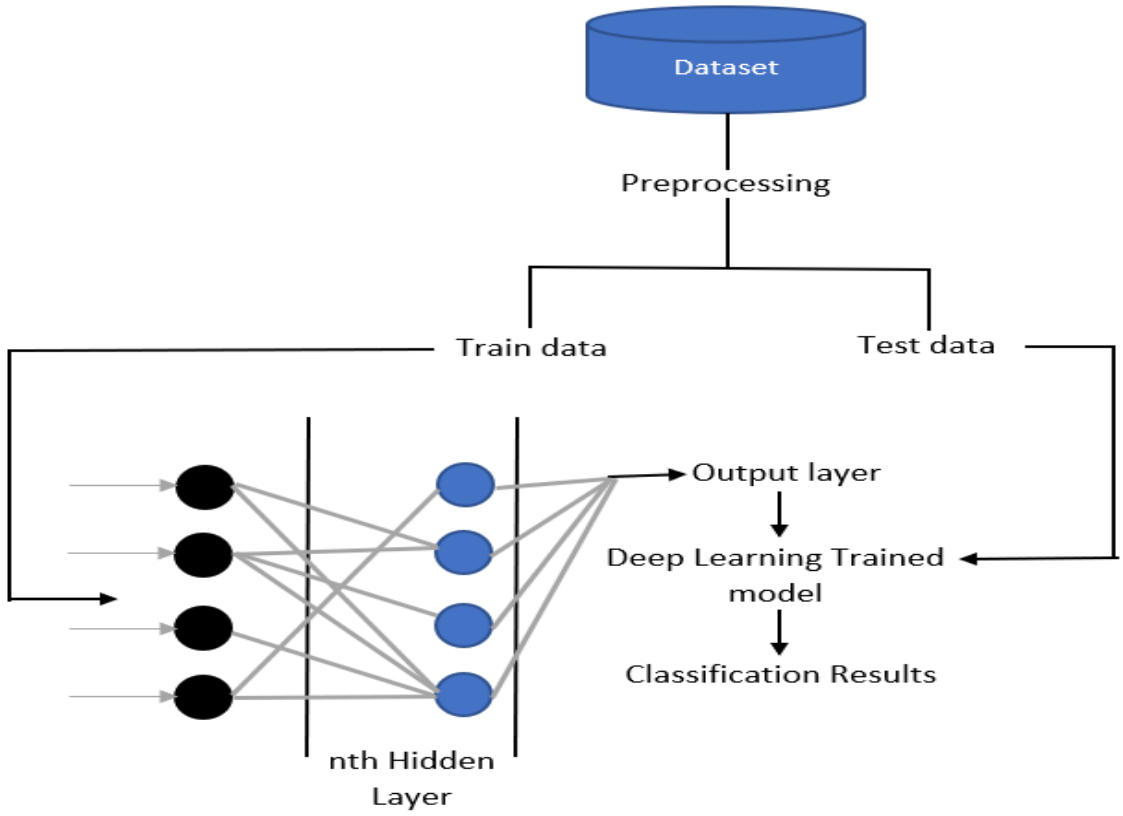

Figure 1: Block diagram of proposed neural network model

This system is contributed in different aspects:

1. A multiple layered deep neural network model has been developed for prediction of heart disease.

2. This paper explores the effect of depth of neural network on the accuracy of heart disease prediction by using different activation functions. Generally, it is found that a deep learning model performs better than other machine learning or data mining techniques.

2. Related Theory

The main goal is to supply services of human related to health management. It helps to modify existing medical system and generate advanced prediction system based on deep neural network which accommodate in development of patient disease, medication, diagnosis, medical regulation. DNN help to expand exploration of medication. Heart disease prediction system includes 13 attributes which are the subset of the attributes from the Cleveland heart disease dataset [11].

Description about dataset:

In this research work, Cleveland heart disease dataset from UCI repository is used. This dataset shows 303 instances amongst which 297 instances contains no null values, while remaining six have some missing attributes. The used dataset have total 76 features in the original form. But many of the published works referred only 13 features, which are shown in Table-1[5].

Table 1 Dataset Table

Feature No.	Feature Description
1	Thallium Scan
2	Sex
3	Number of Major Vessels Colored by Fluoroscopy
4	Maximum Heart Rate achieved
5	Serum Cholesterol
6	Old Peak
7	Peak Exercise Slope
8	Fasting Blood Sugar
9	Exercise Induced Angina
10	Age(Years)
11	Resting Blood Pressure
12	Resting Electrocardiographic Results
13	Chest Pain Type

Providing Front End for Deep Learning Model:

Front end for machine learning model was designed by using python libraries such as flask [12]. In the proposed system we have used flask for providing front end connection. Flask web framework for deep learning model was used. HTML5 is used for GUI and flask is server for connecting deep learning model to the GUI and providing inputs to the deep learning model.

3. Literature Review

Liaqat Ali [5] designed a system which predicts the heart disease based on deep neural network with a statistical model that is a chi^2 model and has proved that the deep neural network is better than the others with respect to the accuracy. The model was built by utilizing dataset with 13 attributes which are stated earlier. The final attributes taken for the training of the model were taken with the help of the chi^2 statistical factors. The training accuracy was 84.05% and testing accuracy was 93.33%. By observing the outcomes, it is clear that the deep neural network along with chi^2 statistical factor gives high accuracy as compared to other algorithms and data mining techniques [5].

The authors in this paper [1] suggested a system to predict the probability of heart attack using machine learning technology with the study of heart signals. The ECG signals are given to systems that carry out a preliminary filtering, and then it make use of a Gustafon-Kessel fuzzy clustering algorithm to apply for signal organization and correlation. The classification identifies the heart diseases such as angina, myocardial infraction and coronary artery diseases.

The authors [2] in this paper proposed a prototype with the datamining techniques in the main Naïve Bayes and WAC (Weighted Associated Classifier). The dataset consists of features such as diabetic, age, sex, height, weight, blood pressure, fasting blood pressure, cholesterol, and hypertension. The system detects whether a patient have heart disease or not. The system is highly scalable and makes probabilistic predictions.

In this paper [3], authors suggest a privileged method for predicting heart disease using two different methods, Naïve Bayes and Logistic Regression. It solves the privacy violation problem with higher accuracy. But it requires more speed and time while training and testing and in this approaches in case of discrete data it leads to some other problems.

Authors in paper [4] deliberated that the medical practitioner's examination plays a remarkable role in the ongoing trend. The findings and investigation of medical practitioner's is the most important concern in real time scenario, as the requirement of training samples and sufficient data's makes these processes more complex. This medical data analysis can be implemented by data mining techniques. It uses the method of SVM (support vector machine). It maximizes the prediction accuracy and avoids over-fitting problem. It classifies and diagnosis the diseases effectively.

In this paper [6] authors developed an intelligent data mining system based on genetic algorithm. To transform data into useful form, encoding was done between range [-1, 1]. The neural network weight optimization by genetic algorithm system uses back-propagation algorithm. The drawbacks were removed in this paper by optimizing the weights of neural network. In this a genetic algorithm specialized for global searching was used.

4. System Architecture

This section describes the deep learning model. The deep learning is used for decision support and prediction and UI is used for data collection. Data collection is done in three types such as at first type personal data like age and sex, at second type, periodic data like sugar and cholesterol. At third type is live data, blood pressure, ECG and heart rating of heart. The input to the website is given manually. The inputs are name, age, sex, cholesterol, diabetic, maximum heart rate achieved and ECG.

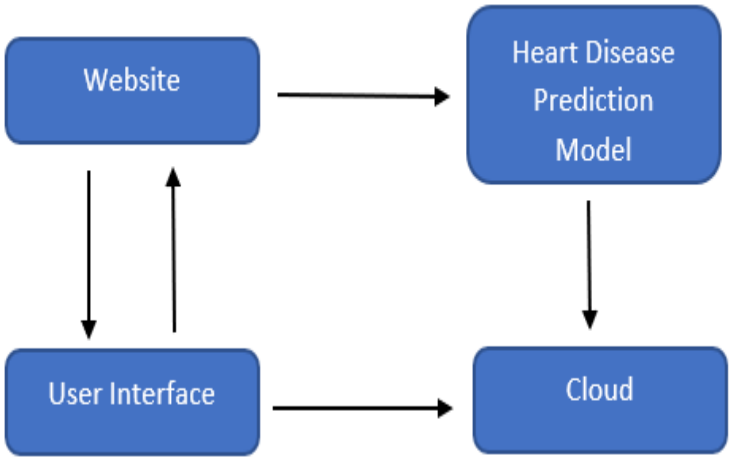

Figure 2 System Architecture of Heart Disease Prediction System

Software's and Libraries Used:

- Scikit Learn
- Tensor Flow
- Keras
- NumPy, pandas
- Flask

Performance Requirements:

To be precise and very specific there are no certain guidelines and benchmarks defined for the web application related to their Performance.

- Reliable and accessible website.
- Validation errors must be displayed on screen.
- Web pages must be interactive and high performance.
- The model with the highest accuracy needs to be chosen for the web application. Since it was observed that the Deep Neural Network was the more accurate with an accuracy of 95.25%, it was chosen.

5. Implementation:

The foremost reason of a predictive model of machine learning is to produce the hypothesis with the help of a various learning algorithm on the training data. The model learns a fitting function by examining the performance of the training data. The hypothesis is produced by reducing errors attained on the training data. The paper gives comparative study of the accuracies obtained by evaluating the trained models on the Random Forest algorithm and the deep neural network model.

Neural Network Model:

The functioning of the neural network is subject to the scope of a network and scope of a network is associated with the count of parameters used in the proposition. If the model has more number of parameters than the required then it leads to the overfitting of the model and in case if the numbers of parameters are less than the required then it leads to the underfitting of the model. Hence a model with ideal number of parameters is required for the best results. To connect the various parameters of the neural network, it is necessary to interpret the assembling the neural network [5]. The neural network works as follows:

Each neuron in a layer takes some inputs, applies some process on it, and then generates an output [5]. The neurons are connected to each other and every connection has a weight. These weights adjusted during the learning phase. They are adjusted such that it gives correct class label to the inputs.

Neural Network Layers:

To perform heart disease prediction using the neural network the below specified neural network model is designed.

- The neural network model contains total eight layers.
- Five layers of which are densely connected layers with an activation function.
- The activation function used in the proposed neural network model is the relu function for the input layer.
- The activation function for the three hidden layers is the Tanh function.
- The activation function for the output layer is the sigmoid function.
- The remaining three layers of the neural network model are the dropout layers.
- The densely connected layers are also known as the fully connected layers. In this every neuron is connected to each and every neuron in the preceding layer.
- The adam optimizer is used in this model to reduce the losses and to gain proper weights and learning rate.

Algorithm of the proposed model:

Input: - $X = a^{[0]}$
$W^{[1]}, b^{[1]}, W^{[2]}, b^{[2]}, W^{[3]}, b^{[3]}, W^{[4]}, b^{[4]}, W^{[5]}, b^{[5]}$, y = actual value
$a^{[5]} = \hat{y}$ = predicted value
$Z = WX + b$
$a = F(Z)$

Z = input to the layer
W = weight of the neuron
X = Input feature vector
b = Bias
a = output of the layer
F = Activation function

Forward Propogation:

$Z^{[1]} = W^{[1]} a^{[0]} + b^{[1]}$
$a^{[1]} = relu(Z^{[1]})$
$Z^{[2]} = W^{[2]} a^{[1]} + b^{[2]}$

$a^{[2]} = \tanh(Z^{[2]})$
$Z^{[3]} = W^{[3]} a^{[2]} + b^{[3]}$
$a^{[3]} = \tanh(Z^{[3]})$
$Z^{[4]} = W^{[4]} a^{[3]} + b^{[4]}$
$a^{[4]} = \tanh(Z^{[4]})$
$Z^{[5]} = W^{[5]} a^{[4]} + b^{[5]}$
$a^{[5]} = \text{sigmoid}(Z^{[5]})$

Backward Propogation:

$\partial a^{[5]} = - [(y/a^{[5]}) – ((1-y)/(1-a^{[5]}))]$
$\partial z^{[5]} = \partial a^{[[5]} (1-(a^{[5]})^2)$
$\partial w^{[5]} = \partial z^{[5]} * a^{[4]}$
$\partial b^{[5]} = \partial z^{[5]}$
$\partial a^{[4]} = \partial z^{[5]} * W^{[5]}$
$\partial z^{[4]} = \partial a^{[[4]} (1-(a^{[4]})^2)$
$\partial w^{[4]} = \partial z^{[4]} * a^{[3]}$
$\partial b^{[4]} = \partial z^{[4]}$
$\partial a^{[3]} = \partial z^{[4]} * W^{[4]}$
$\partial z^{[3]} = \partial a^{[[3]} (1-(a^{[3]})^2)$
$\partial w^{[3]} = \partial z^{[3]} * a^{[2]}$
$\partial b^{[3]} = \partial z^{[3]}$
$\partial a^{[2]} = \partial z^{[3]} * W^{[3]}$
$\partial z^{[2]} = \partial a^{[[2]} (1-(a^{[2]})^2)$
$\partial w^{[2]} = \partial z^{[2]} * a^{[1]}$
$\partial b^{[2]} = \partial z^{[2]}$
$\partial a^{[1]} = \partial z^{[2]} * W^{[2]}$
$\partial z^{[1]} = \partial a^{[[1]} (1-(a^{[1]})^2)$
$\partial w^{[1]} = \partial z^{[1]} * a^{[0]}$
$\partial b^{[1]} = \partial z^{[1]}$

Loss Function:

In this proposed model the loss function used is the mean squared logarithmic loss. As the data may have spread values and while predicting the model should not be punished heavily.

$L(y, \hat{y}) = (1/N) \sum_{i=0}^{N} (\log(y_i+1) – \log(\hat{y}_i+1))^2$

Random Forest Algorithm:

In machine learning random forest is a supervised machine learning algorithm that creates the forest of decision trees. For example, if blood pressure is above a specific value then stage1 is output and if blood pressure is below specific values then stage 0 is output and by combining these two-decision tresses it calculates the final output. The forest it builds is an assembly of Decision Trees, most of the time trained with the “bagging” method. Algorithm creates multiple decision trees and collects the results of each decision tree together to get a more accurate and stable prediction [10].

Random Forest pseudocode:

I. Haphazardly select "a=2" attributes from total **"**t=13**"** attributes. Selected attributes are blood sugar, cholesterol, age, sex, ECG, maximum heart rate and chest pain should be less than t.

II. From the "13" attributes, calculate the best attribute "d", node d is root of decision tree using the best point for split.

III. Split root into subsets using best split point.

IV. Continue the process of above three steps until a single attribute is found.

V. Construct forest of decision trees by reiterating steps I to IV for creating as many as 'n' number of trees.

Heart Disease Prediction Pseudocode using neural network trained model:

To perform heart disease prediction using the trained neural network model uses the below pseudocode.

I. Inputs the heart disease prediction attributes like blood sugar, cholesterol, age, sex, ECG, maximum heart rate, chest pain and other. Then apply the model of trained neural network for the purpose of prediction (i.e. the output such as stage 0, stage 1, stage 2 and stage 3, stage 4 and saves the predicted output)

II. Calculate the probability for each predicted output like stage 0 to stage 4.

III. Recognize the maximum probability as the predicted outcome for heart disease prediction using neural network trained model.

6. Result and Discussion

In this paper, heart disease prediction was the diagnosed in form of Normal, Stage1, Stage2, Stage3, Stage4 (critical) and health tips were provided accordingly. The normal condition shows that the person do not have any heart related issues, all the 13 parameters are in the normal range. The stage 1 shows the start of the artery blockage with its related tips. Similarly the further stages show the increase in the artery blockage with that stage related tips which should be followed. The stage four shows the high artery blockage and advice immediate surgery with the concern of the surgeon. Following are the snapshots of the website.

Below is the snapshot of the home page. On this page user has to enter the data that is required to analyse and predict the cardiac dieses.

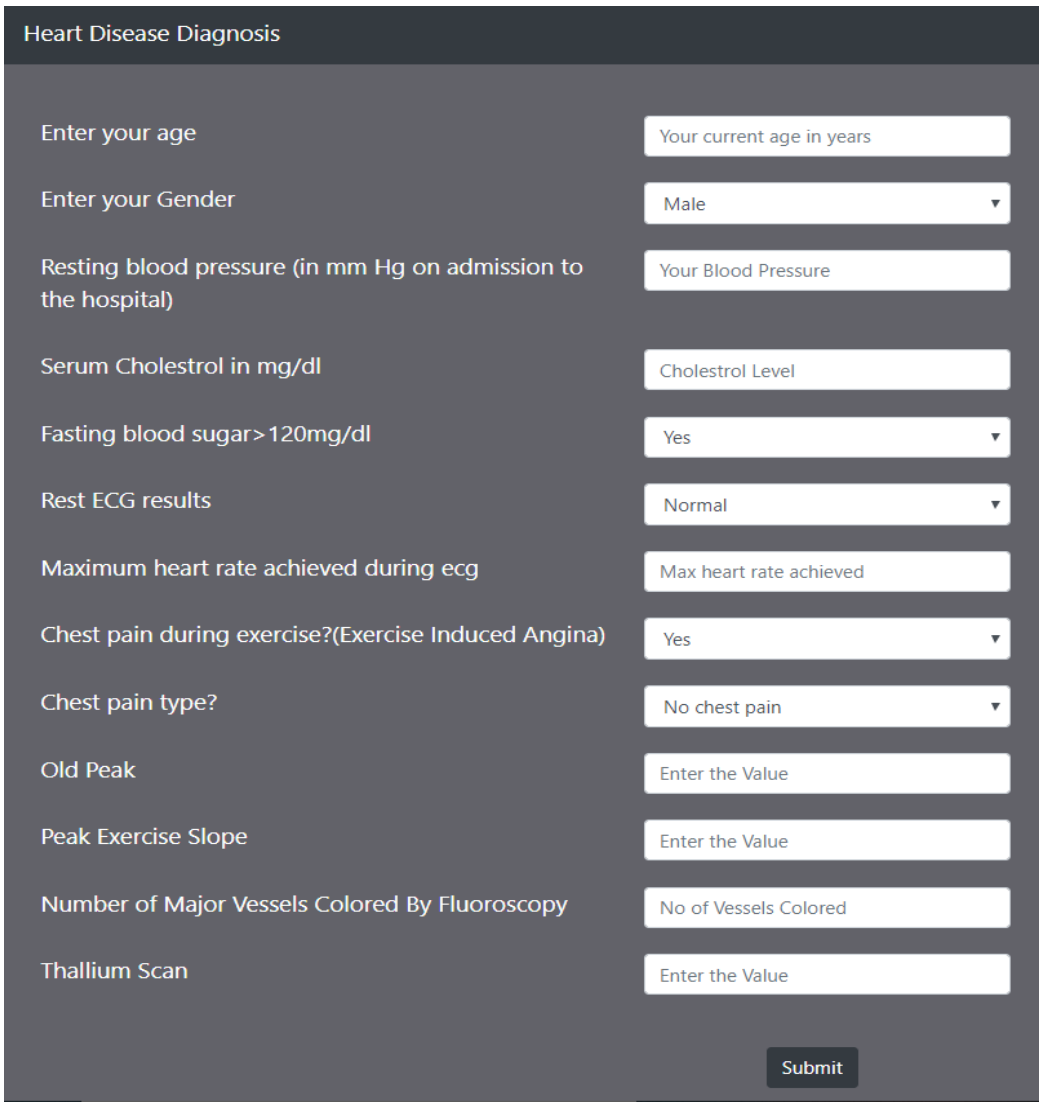

Figure 3 Personal and cardiac paramemters entry page

In Figure 4, prediction of heart disease and health tips are shown.
The prediction of heart disease is in terms of stages from normal to critical have been diagnosed. And health tips are generated according to the stages. Stage 1 diagnosis patient can adopt the tips like exercise regularly, quit smoking, treat high blood pressure, treat high cholesterol, discontinue alcohol and drugs consumption and avoid the further health complications.

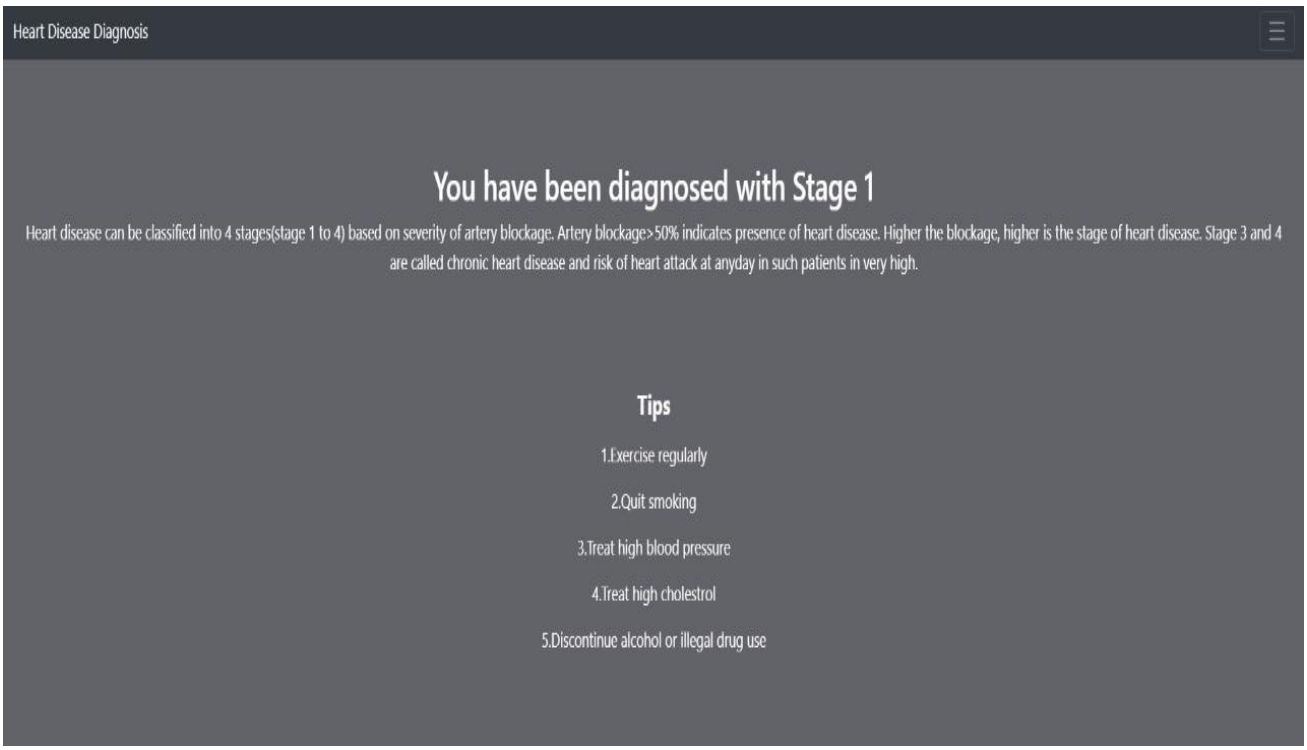

Figure 4 Heart Disease Prediction and health tips

Figure 5 describes the accuracy of the neural network and it is 95.25%. The accuracy is obtained on the test data with 76 instances.

```
loss = model1.evaluate(X_test, Y_test, verbose=1, batch_size=30)
```

```
76/76 [==============================] - 0s 1ms/step
```

```
loss
```

```
0.04749996180793172
```

```
print("Final accuracy is {}".format(100-loss*100))
```

```
Final accuracy is 95.25000381920682
```

Figure 5 Accuracy of neural network model

Figure 6 shows the graph of the training and validation loss of the deep neural network with dense layers. The loss shown with the blue line is the training loss of the model and the loss shown with the orange line is the validation loss.

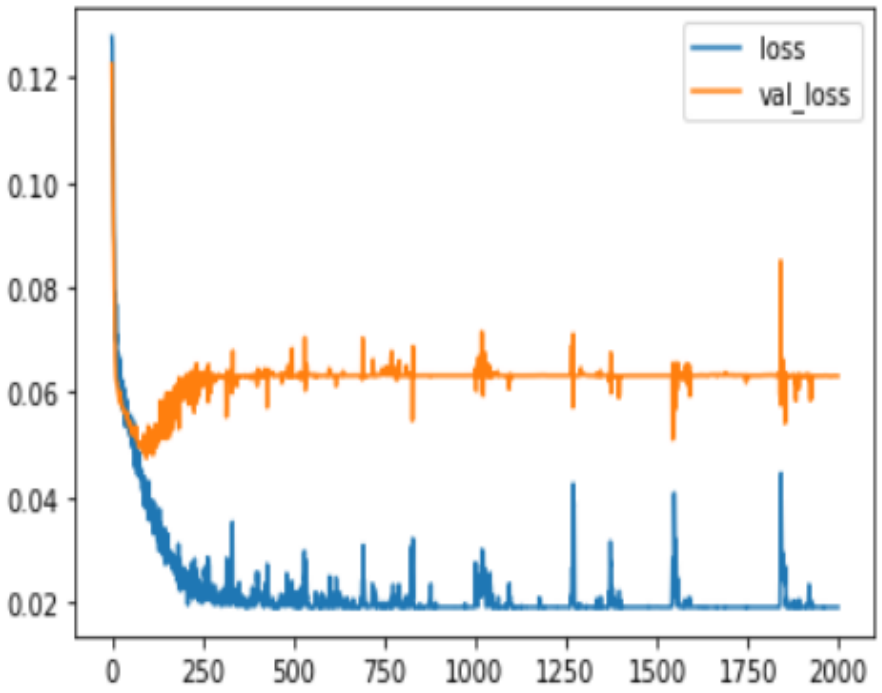

Figure 6 Graph of loss of the neural network model

Figure 7 shows the accuracy of the machine learning model is 81.57% on Random Forest algorithm.

```
The Accuracy of Random Forest Algorithm is: 81.57894736842105

Confusion Matrix:  [[25  6]
 [ 8 37]]
```

Figure 7 Accuracy of Random Forest Algorithm

The table 2 shows the comparisons of the accuracies of the various classification methods used for heart disease prediction. The data mining technique, random forest gives result by implementing multiple decision trees; it gives good result on small data, but as per our work shown in below table the results obtained is not up to the mark. The support vector machine (SVM) gives good result on complex data, but the result observed in this are also not up to the mark. The χ^2 Statistical Model by Liaqat Ali [5] takes 9 features amongst 13 using the χ^2 statistical method of feature extraction. In his research he have obtained a good result with the K-fold cross validation and a better using the Holdout cross validation.

In our research work we have obtained better result than the χ^2 statistical model of DNN with holdout cross validation by Liaqat Ali [5]. In our work we obtained the accuracy of 95.25% by taking all the 13 features and by hyper parameter tuning.

Table 2 Classification accuracies of some different models

Method	Accuracy (%)
Random Forest Model	81.57
SVM Model	85.52
χ^2 Statistical Model + DNN Liaqat Ali [5]	91.57 (K-fold)
χ^2 Statistical Model + DNN Liaqat Ali[5]	93.33 (Holdout)
Deep Neural Network Model with Tanh function in the hidden layers (Proposed)	**95.25**

7. Conclusion

The Random Forest algorithm and deep neural network model were implemented. Datasets were trained for both, individually. After this, all of them were tested. The most efficient model was to be selected based on various criteria. We found out that deep neural network model was the most efficient out of the two with an accuracy of 95.25%. Random Forest had accuracy of 81.57%. Thus, deep neural network model was further implemented using a better user interface in form of a web application. This would help the end users get a preliminary prediction about the condition of their heart. Since heart diseases are a major killer in India and throughout the world, application of a promising technology like deep learning to the initial prediction of heart diseases will have a profound impact on the society. This will tell the user if they are at a risk and if they need to take any action in the critical situations. This will help reduce the death rate due to the heart attacks.

Hence by using the above approach successfully, analysis of heart diseases of the individual was performed and the result was obtained which predicted the risk of heart disease based on the parameters provided by the user. In future it can be extended as the wearable device as an assembly of all hardware devices and can be accessible via mobile application. More trained dataset is required to get the highest accuracy for prediction of cardiac problems.

References

[1] de Carvalho Junior, Helton Hugo, et al. “A Heart disease recognition embedded system with fuzzy cluster algorithm.” Computer methods and programs in biomedicine 110.3 (2013): 447-454.

[2] Patel, Ajad, Sonali Gandhi, Swetha Shetty and Bhanu Tekwani. “Heart Disease Prediction Using Data Mining.” (2017)

[3] Waghmode, Mr. Amol A., Mr. Darpan Sawant and Devenb D. Ketkar. “Heart Disease Prediction Using Data Mining Techniques.” Heart Disease (2017)

[4] Senthil Kumar, B., and Dr Gunavathi R. “A Survey on Data Mining Approaches to Diabetes Disease Diagnosis and Prognosis.” IJARCCE 5 (2016):463-467.

[5] Liaqat Ali, Atiqur Rahman, Aurangzeb Khan, Mingyi Zhou, Ashir Javeed and Javed Ali Khan “An Automated Diagnostic System for Heart Disease Prediction based on χ^2 Statistical Model and Optimally Configured Deep Neural Network. IEEE (2019):34938-34945.

[6] Amin, S. U., Agarwal, K., Beg, R.: Genetic neural network based data mining in prediction of heart disease using risk factors. In: IEEE Conference of Information and Communication Technologies (2013).

[7] Z. Arabasadi, R. Alizadehsani, M. Roshanzamir, H. Moosaei, and A. A. Yarifard, "Computer aided decision making for heart disease detection using hybrid neural network-genetic algorithm," Comput. Methods Programs Biomed. , (2017).

[8] H. Yan, Y. Jiang, J. Zheng, C. Peng, and Q. Li, "A multilayer perceptron based medical decision support system for heart disease diagnosis," Expert Syst. Appl., (2006).

[9] K. Vanisree and J. Singaraju, "Decision support system for congenital heart disease diagnosis based on signs and symptoms using neural networks," Int. J. Comput. Appl., (2011).

[10] Algorithm Used:
https://towardsdatascience.com/the-random-forest-algorithm-d457d499ffcd

[11] Dataset
https://archive.ics.uci.edu/ml/datasets/heart+Disease

[12] Flask
https://www.toptal.com/python/python-machine-learning-flask-example