ANN BASED HEART DISEASE PREDICTION SYSTEM

[1]Shreya Agrawal, [2]Malay P Doshi, [3]Urvi Kanade, [4]Devaki Upadhyay, Prof. K Jayamalini
[1]Author, [2]Author, [3]Author, [4]Author
Department of Computer Engineering
Mumbai University
Shree LR Tiwari College of Engineering, Thane, India

Abstract: Artificial Intelligence techniques have been successfully employed in disease diagnosis, disease risk evaluation, patient monitoring, robotic handling of surgeries and predicting effect of new medicines. This paper proposes and evaluates Neural Networks for diagnosing Hypertension risk. Risk factors viz. Systolic and Diastolic Blood Pressure, Body Mass Index, Heart Rate, Cholesterol, Glucose, Stress, Diet have been taken as inputs to the system. The system classifies the input samples and predicts whether heart disease is present or absent. The results of proposed system have been compared with Neural Network in terms of Accuracy, Mean Square Error and Regression and found better. This qualitative approach is then integrated with genetic algorithm to diagnose the presence of the disease.

Index Terms - **Neural network System, Genetic Algorithm, Optimization of Back Propagation Network, Heart Disease Prediction Model, Layers, Neurons.**

I. INTRODUCTION

The current scenario for the diagnosis of heart diseases uses clinical dataset having parameters and inputs from complex tests conducted in labs. None of the system predicts heart diseases based on risk factors such as age, family history, diabetes, hypertension, high cholesterol, tobacco smoking, alcohol intake, obesity or physical inactivity, etc. heart disease is one of the leading causes of deaths all over the world. The World Health Statistics 2012 report enlightens the fact that one in three adults worldwide has raised blood pressure – a condition that causes around half of all deaths from stroke and heart disease. Heart disease, also known as cardiovascular disease (CVD), encloses a number of conditions that influence the heart – not just heart attacks. Heart disease was the major cause of casualties in the different countries including India. Heart disease kills one person every 34 seconds in the United States. In addition, if at all someone wants to make sure he/she is not suffering from any kind of heart related problem then there are a lot of expensive clinical tests are involved which not everyone can afford. Heart disease patients have lot of these visible risk factors in common which can be used very effectively for diagnosis. System based on such risk factors would not only help medical professionals but it would give patients a warning about the probable presence of heart disease even before he visits a hospital or goes for costly medical check-ups

This project is a computer-based clinical decision support and can reduce medical errors, improve patient safety and reduce unnecessary changes in practice, and improve the prognosis of the patient's medical history to integrate patients. This proposal is promising as data modelling and analysis tools, such as data mining, you must generate a rich environment or knowledge, can help to significantly improve the quality of clinical decision-making potential. The main objective of this study was to develop a prototype of heart disease forecasting system using data mining and neural network concepts. By providing effective treatment, but also help to reduce the cost of treatment and improve the visualization and ease of explanation, a huge knowledge and accurate data in the field. Big companies invest heavily in this type of activity in order to focus on the risks involved and possible events. This work brought on as a reasonable expectation for the future development of the basis for all previous and current data together. This technique involves two most successful data mining tools, neural networks and genetic algorithms. The proposed system will be implemented as a desktop application, where user will give answers to the predefined questions. The system will retrieve the data from stored database collection and compares the user values with trained data set using Multilayer perceptron neural network. Back propagation algorithm will be used to train the network using the weights optimized by
Genetic algorithm.

II. RELATED WORK

[1] D. K. RAVISH, NAYANA R. SHENOY, DR. K. J. SHANTHI, S. NISARGH paper is based on how ECG (Electrocardiography) data clubbed with certain clinical parameters can be used to predict heart disease. It made use of various algorithms that is, Artificial Neural Network (ANN), Back Propagation Training Algorithm(BPT) and various other parameters like layers, weights and biases of ANN. [2] THERESA PRINCY R., J. THOMAS paper is based on case study between various algorithms and evaluation of various algorithms for predicting Heart Disease.[3] C. KALAISELVI paper is based on Average KNN (K- Nearest Neighbor) algorithm for predicting Heart Disease.[4]AIGERIM

ALTAYEVA, SULEIMENOV ZHARAS, YOUNG IM CHO paper is based on integrating K-means and Naïve Bayesian algorithm for predicting Heart Disease. **EXISTING SYSTEM:**

(A) Artificial Neural Network

An artificial neural network is an interconnected group of nodes in the network which stimulates organization of neurons in a brain. Here, each circular node represents an artificial neuron and an arrow represents a connection from the output of one artificial neuron to the input of another. Such systems "learn" to perform tasks by considering examples, generally without being programmed with any task-specific rules. An ANN is based on a collection of connected units or nodes called artificial neurons which loosely model the neurons in a biological brain. Each connection, like the synapses in a biological brain, can transmit a signal from one artificial neuron to another. An artificial neuron that receives a signal can process it and then signal additional artificial neurons connected to it. The input data is passed to neuron in an input layer and processing is performed in hidden layer and then output is produced in output layer using any of the ANN algorithms. The ANN architecture comprised of 100 Neuronal layers.

(B)Back Propagation Training Algorithm

When a training input-output example is presented to the system, the back-propagation algorithm computes the system output and compares it with the desired output of the training example. The error is propagated backwards through the network from the output layer to the input layer. The neuron activation functions are modified as the error is propagated. To determine the necessary modifications, the back-propagation algorithm differentiates the activation functions of the neurons.

For each example e in the training set do

1. O = neural-net output(network,e) ; forward pass

2. T = teacher output
   - Calculate error (T - O) at the output units
   - Compute delta_wi for all weights from hidden layer to output layer; backward pass
   - Compute delta_wi for all weights from input layer to hidden layer; backward pass continued.

3. Update the weights in the network end until all examples classified correctly or stopping criterion satisfied

4. return(network).

This algorithm aims to reduce the differences in output for predicting heart disease by comparing the result with trained data and then if there is difference between 2 values, an iterative process of weight adjustment goes on unless the correct result for indicating presence of heart disease is predicted

(C) K- Nearest Neighbor

The k nearest training set data are found for each row of testing data and the classification is done by determining the majority vote by breaking the ties at random. Each neighbor is assigned a weight such that the nearest neighbors contributes high to the average than the distant neighbors.

Disadvantages of existing systems:

Most of the systems use hard coded weight and biases for computing output layer value, thus the result produced is not optimized.

Systems are based on Cleveland Heart Disease dataset which comprises of complex attributes which are not easily available as patients need to go undergo through complex and expensive tests. It attempts to mimic the result as indicated by angiography process. The cost of angiography test is less as compared to the bundled-up costs of complex and expensive tests. So, such systems are not used much.

With increase in complexity of attributes, people cannot use the existing system.

Efficiency of KNN depends on number of clusters chosen.

KNN works only when data is cleaned and normalized. It is incapable of tolerating any noise

III PROPOSED SYSTEM:

The objective of this research is to create an intelligent & cost-effective system which will overcome the limitations of existing system and improve its performance. A novel way to enhance the performance of a model that combines genetic algorithms and ANN for feature selection and classification is proposed. Thus, a reliable method for both feature selection & classification is required. The feature selection is based on a new genetic algorithm and classification is based on Artificial Neural Network (ANN). Feature selection is another factor that impacts classification accuracy. By extracting as much information as possible from a given data set while using the smallest number of features, we can save significant computation time and build models that generalize better for unseen data points. This document presents a forecast methodology using Artificial Neural Network

(ANN) and Genetic Algorithms (GA). The GA has been used for the selection of the training inputs of the ANN in order to minimize the training result error.

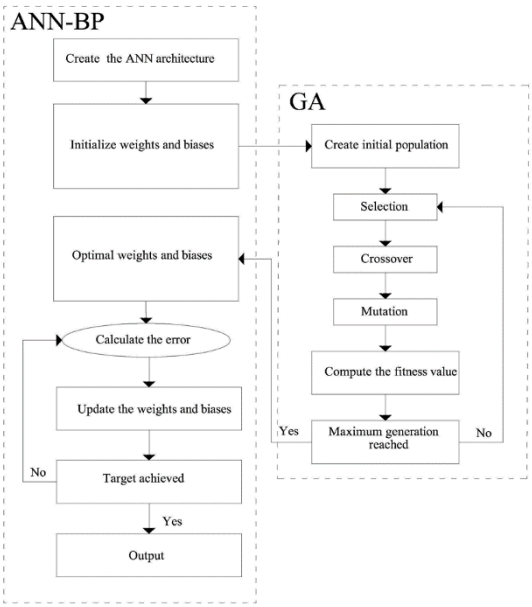

fig.1 Proposed Block diagram using ANN-BP with GA

3.1) POPULATION AND SAMPLE:

For accurate and real time results prediction it was decided to consider patients into sample population to predict the heart disease. A survey was conducted by group members and the responses of patients were recorded in the registers.

3.2) DATA AND SOURCES OF DATA:

American Heart Association (AHA) has stated that basic parameters like age, gender, diabetes, cholesterol and so on are vital parameters to predict the presence of Heart Disease. 302 records were collected from patients at Medways Cardiac clinic under the guidance of cardiologist Dr. Harminder Singh. The patients were asked basic questions about like their age, weight, lifestyle and how their heart is functioning in order to relate these factors and condition of heart.

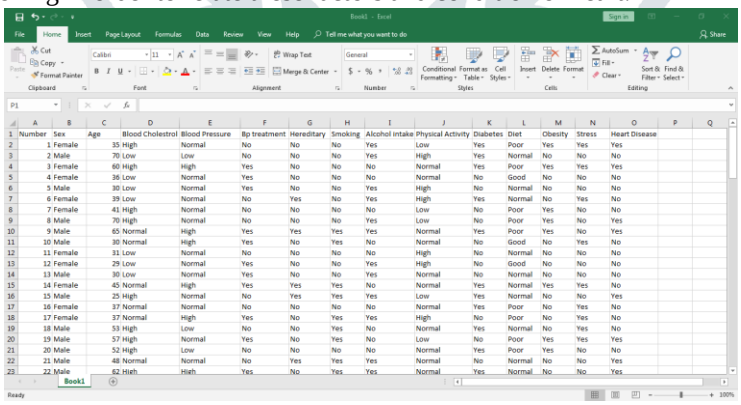

fig.2 Screenshot of sample data collected

3.3) WORK-FLOW:

- **LOGIN(ADMIN):** This Module is the entry point for the admin. It is responsible to give access of the system to intended users. Admin has the power to control and monitor the system. The admin is responsible to perform training of the system as well.

- **NETWORK CREATION:** This model is the next phase after the admin has given access to system. In this model, parameters are specified which can be used for creating Artificial Neural Network like type of network, number of layers and so on.

- **TRAINING MODEL:** After the network is created, the model needs to be trained for predicting the results at output layer. This is the crucial part of the system as the accuracy of system is dependent on training where training is performed on records collected and confusion matrix is plotted.

- **LOGIN(DOCTOR):** After the above-mentioned processes are completed, the next phase is the doctor's login. Since this system handles an enormous amount of data which goes on increasing, this module is added so that patient cannot misuse this system.

- **USER REGISTERATION:** The user registers itself with the system by entering the phone number and name.

- **RESULT PREDICTION:** After the user has registered itself, the user can answer parameters as mentioned on the screen and check the condition of the heart.

- **GRAPH PLOTTING:** Once the system has predicted result, a comparative graph is plotted to show the differences in errors with and without using Genetic Algorithm.

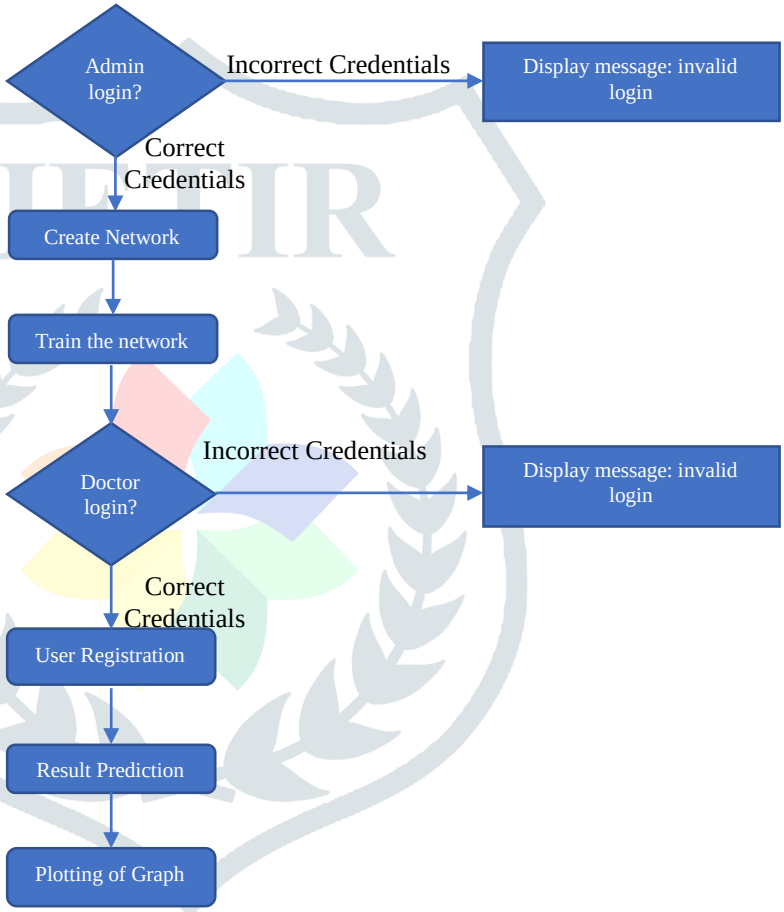

fig. 3 Workflow of System

3.4) ADVANTAGES:

Early Prediction of Heart Disease before it becomes a serious issue. The user ignores the symptom such as high blood pressure, obesity, effect of physical activity and so on. These parameters play an important role in monitoring the heart.

User can easily access this system as it requires very basic and common parameters and these parameters are easily understood by the doctor.

It requires very basic blood tests such as diabetes and cholesterol which most patients have gone through during routine check-up.

An attempt is made to check routinely heart monitoring and take actions based on report predicted such as going through tests in depth like ECG, 2D Echo, Stress Testing and so on. This ensures that routine maintenance can be performed at affordable and cheaper rates.

IV. SOFTWAREREQUIREMENTS:

Operating System: Windows 7/8/10

Technology: MATLAB

Database: Ms Excel

MATLAB Version: MATLAB Version 2015b and above

V. RESULTS AND DISCUSSION:

This section deals with the results of the system implemented and are discussed below:

1. The given figure below shows the Create Network window where the user(admin) specifies the parameter to create the network viz. Network Type, Number of Layers, Number of Epochs, Transfer Function, Backpropagation Network Training Function and Performance Function. After clicking on Create Network, it creates the network as parameters specified by the user.

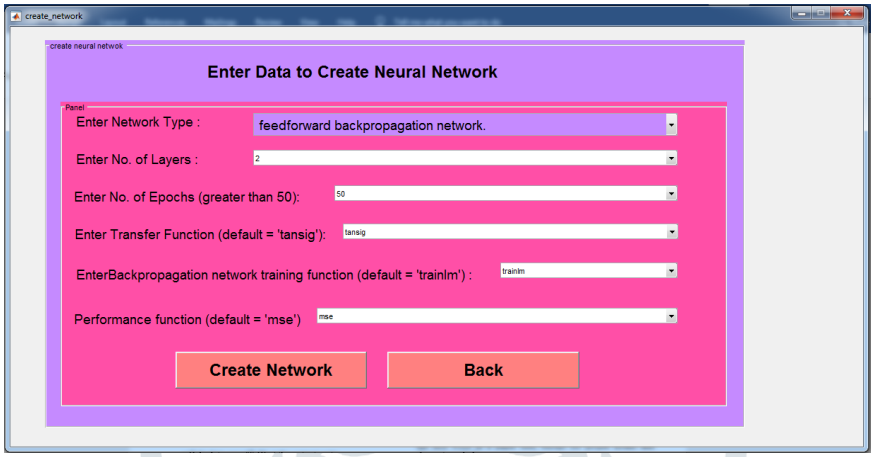

fig. 4 Create Network Screen

Epochs describe the number of cycles required in backpropagation network to obtain accurate result. Transfer function describes activation function which maps particular input to output. The trainlm is a network training function that updates weight and bias values. The mse is a network performance function. It measures the network's performance according to the mean of squared errors.

2. The figure below shows the train network window which consists of following buttons : 1.Start Training- to start the training of the system.2. Test Data- it directs to the main window where the user can enter data. 3. Back- it directs to the previous window.

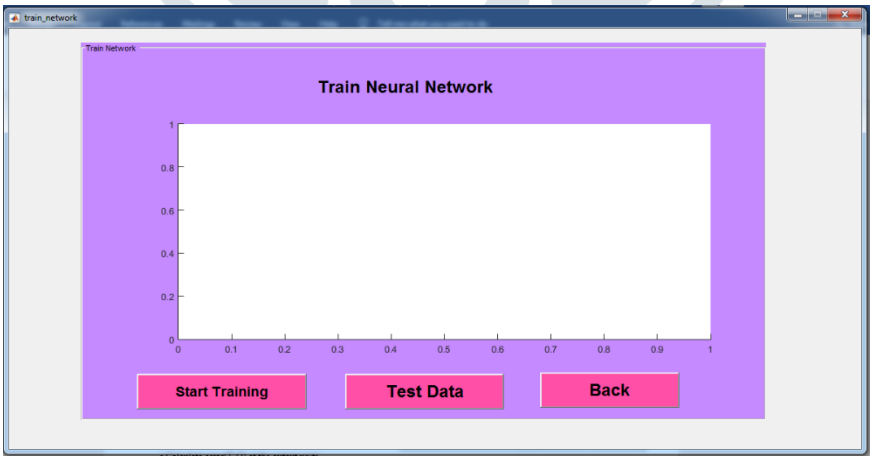

fig. 5 Training Neural Network

This window is responsible for training input dataset using MATLAB's inbuilt training tool nntraintool and plots confusion matrix.

3. The two figures below shows the neural network training (fig. 6) tool and the confusion matrix (fig. 7). The confusion matrix shows the accuracy of the system.

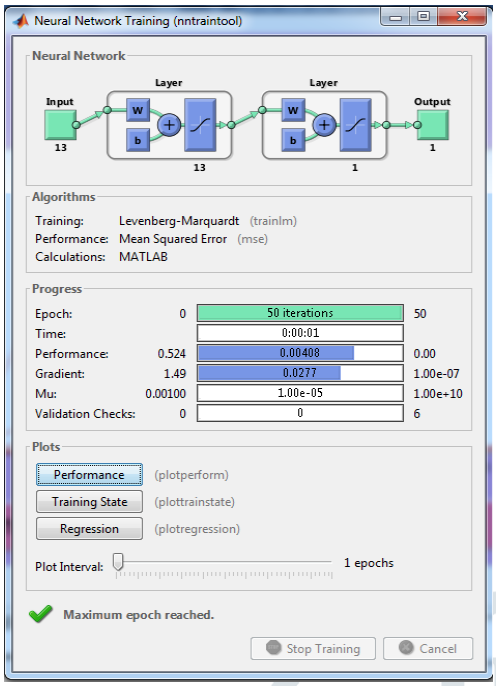

fig. 6 MATLAB's inbuilt training tool

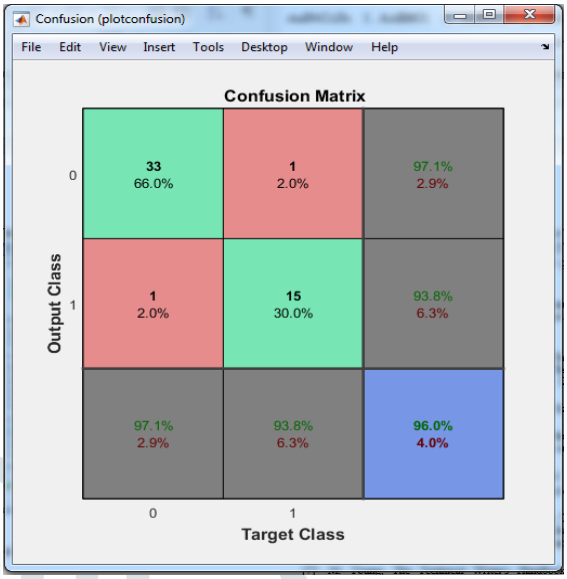

fig. 7 Confusion Matrix

The above 2 figures show training of datasets and confusion matrix with 96% accuracy for 2 layers and 50 epochs as indicated in the figure 6. This window is the heart and soul of the project.

4. The figure below shows the main GUI window with following parameters: 1. Age 2. Sex 3. Blood Cholesterol 4. Blood Pressure 5. Bp treatment 6. Hereditary 7. Smoking 8. Alcohol Intake 9. Physical Activity 10. Diabetes 11. Diet 12. Obesity 13. Stress

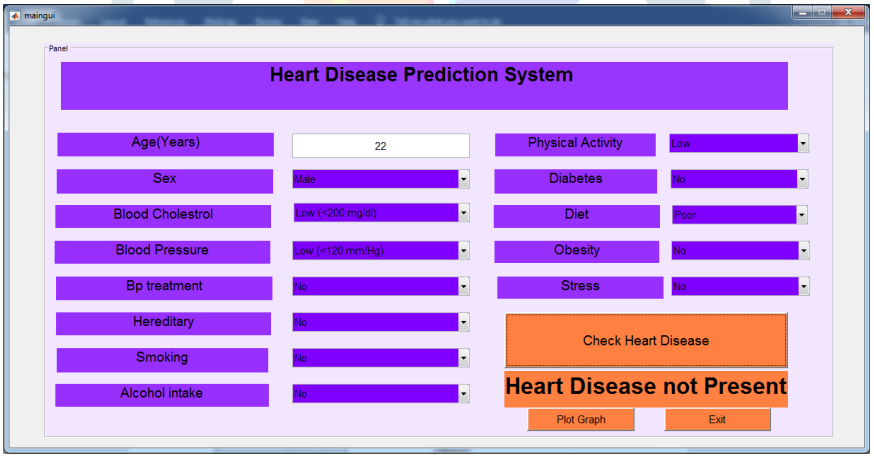

fig. 8 Screen to Enter Patient's details

This screen takes the input parameters from the patients and feeds it to the Neural Network architecture to predict whether heart disease is present or not. This screen is the one accessed by patients where the accuracy of the results depends on the training of dataset.

5. The figure below shows the graphs plotted. Graph 1 is showing errors with and without Genetic Algorithm and Graph 2 is showing Variance in Generations.

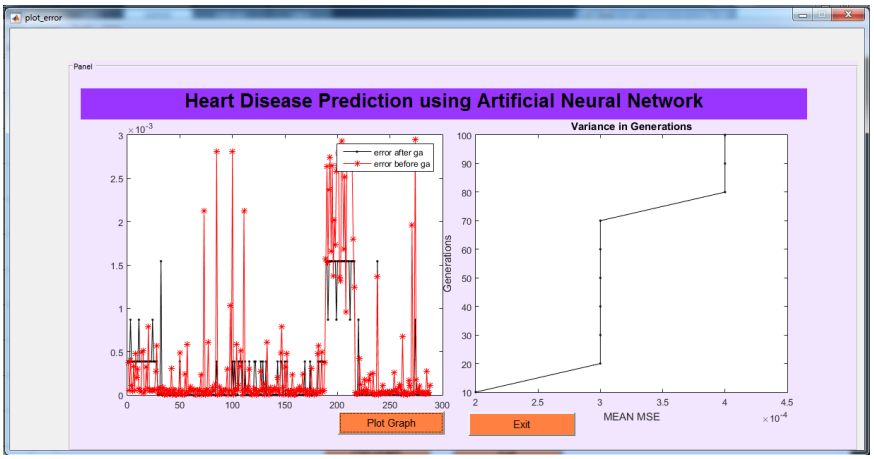

fig. 9 Graph Plotting

This window just gives comparison in the performance in terms of errors that is optimization with GA versus optimization without GA

VI. CONCLUSION & FUTURE SCOPE:

In this paper, A prototype heart disease prediction system will be developed using data mining techniques. The system will extract hidden knowledge from a historical heart disease database. The models will be trained and validated against a test dataset. we apply a combined ANN with GA approach to the heart disease diagnosis problem. The objective of the work is to find the presence of heart disease. The proposed work also helps to minimize the cost and maximize the accuracy. This system has lot of future scope as the need for this system is felt very greatly. The few numbers of future implementations are:

1. The patient can login into system by sending OTP on the phone number of the registered patient.
2. If the report is predicted as 'Heart Disease not present', the borderline values must be highlighted which at present needs urgent action.
3. An attempt to incorporate dietician section in this system will provide the users all thing needed under 1 roof.
4. Finally, the admins can also remove the doctors who misused the system. This can be done through feedback which patients provide.

IV. REFERENCES:

[1] D. K. Ravish, K. J. Shanthi, N. R. Shenoy and S. Nisargh, "Heart function monitoring, prediction and prevention of Heart Attacks: Using Artificial Neural Networks," *2014 International Conference on Contemporary Computing and Informatics (IC3I)*, Mysore, 2014, pp. 1-6. doi: 10.1109/IC3I.2014.

[2] T. Karayılan and Ö. Kılıç, "Prediction of heart disease using neural network," *2017 International Conference on Computer Science and Engineering (UBMK)*, Antalya, 2017, pp. 719-723. doi: 10.1109/UBMK.2017.

[3] S. A. Sabab, M. A. R. Munshi, A. I. Pritom and Shihabuzzaman, "Cardiovascular disease prognosis using effective classification and feature selection technique," *2016 International Conference on Medical Engineering, Health Informatics and Technology (MediTec)*, Dhaka, 2016, pp. 1-6. doi: 10.1109/MEDITEC.2016.7835374

[4] A.Dewan and M. Sharma, "Prediction of heart disease using a hybrid technique in data mining classification," *2015 2nd International Conference on Computing for Sustainable Global Development (INDIACom)*, New Delhi, 2015, pp. 704-706.