IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 6, June 2016

Pre-Prediction of Tuberculosis Disease Using Soft Computing Technique

Rupali Zakhmi[1], Jyoti Arora[2]

Research Scholar, Computer Science and Engineering, Desh Bhagat University, Mandi Gobindgarh, India[1]

Assistant Professor, Computer Science and Engineering, Desh Bhagat University, Mandi Gobindgarh, India[2]

Abstract: Tuberculosis is an epidemic disease that points to death. It spreads through the air by person anguish from Tuberculosis. Ten or more people can be infected per year by a single patient. Like any transmittable disease, TB is ubiquitous even in urbanized nations. But it is a foremost problem in the established nations including India. Soft computing is an agglomeration of methodologies. For tuberculosis detection, consolidation of the Genetic algorithm and Neural Network is proposed. This soft computing technique is used to acquire an optimized solution that is needed to seize correct decision for improved performance and gave the rigorous result. The main aspiration of this scrutiny is to pre-predict information about tuberculosis disease at low outlay. Genetic Algorithm has Fitness function, Crossover and Mutation operations to disentangle optimization problems. In this study, various parameters are taken as an input and conform five hidden nodes afterward hatch an output. The Neural Network is trained with Genetic Algorithm and Back propagation algorithm to improve accuracy. There are ten parameters such as Age, Gender, Cough, Fever, Chest Pain, Weight Loss, Night Sweats, Unwillingness for work and Loss of Appetite. These parameters are adopted by many researchers to identify Tuberculosis.

Keywords: Tuberculosis, Genetic Algorithm, Neural Network, Soft Computing, Data Mining.

I. INTRODUCTION

Tuberculosis is one of the well known disorders for all persons in developing countries including India. It is the extreme communal cause of death in personage. It is a disease triggered by bacteria which strikes human body parts, mainly lungs. Tuberculosis patient's infection gets treated with medicine and cure from this disease is very difficult. It is only achievable by adopting full course of medicine. The physician takes a long time period to recognize tuberculosis.

The objective of this paper is to build a data mining solution which tends to make medical diagnosis of tuberculosis rigorous and will abet choosing if it is reasonable to start tuberculosis treatment on alleged victims without waiting the particular medical test outcomes or not. Based on parameters check whether person has disease or not. There are various parameters which are used for predicting tuberculosis using soft computing.

Disease Prediction is a prime issue in data mining. There are a lot of diseases that are predicted using diverse data mining techniques such as Breast cancer, Tuberculosis, Heart Disease, Liver and Diabetes Prediction [11]. Some researcher uses the amalgamation of Support vector machine and Decision tree to diagnose Breast Cancer [8]. To improve classification performance, C 4.5 Decision Tree and SVM classifier was introduced with an accuracy of 96.9 %. Navies Bayesian and Weighted Associative Classifier are used to predict diabetes and heart disease. Doctor employs these data mining techniques to check whether a person will suffer from this disease or not in the future [9]. On the other hand, Liver disease is predicted using SVM and Naive Bayes Algorithm [7]. In proposed work, Tuberculosis Disease is predicted using soft computing technique such as Hybrid Genetic Algorithm and Neural Network.

1.1 GENETIC ALGORITHM

The genetic algorithm (GA) is a catechistic algorithm used to spawn valuable outcome for optimization problems. It produces solutions to problems using diverse techniques such as fitness and crossover and mutation [5].

The genetic algorithm commences with the random creation of a set of individuals termed as population. The solution obtained from one population is used to create a new population. A recently generated population is compared with the earlier one and annihilate the worst population. This process continues until desired upshot attained. New offspring (solutions) are elected by the fitness function. The main stages are-

1. Outset random creation of population.
2. Reckon the fitness function.
3. Replicate the process.
   i. Randomly opt two chromosomes as a parent.
   ii. Apply crossover function on chosen parent chromosomes to beget two offspring.
   iii. Carry out mutation on two offspring and append the resulted value to the population.
   iv. Weigh up the fitness of acquired chromosomes.
4. Assume the optimum chromosome for next generation.
5. Check step 3

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 6, June 2016

Genetic Algorithm is a search approach based on the scheme of evolution. A Hybrid Genetic Algorithm Neural Network is developed by applying Genetic algorithm (GA) on Neural Network. To determine the weight of Neural Network, a genetic algorithm approach is introduced. GA will form points of a promising solution. The primary task is to find out fitness function for getting high-quality results. The basic mode of genetic algorithm is as follow:-

1. Selection – Fetch a pair of parent chromosome from a recent population based on their fitness. To search out the overall fitness, selection assists to eliminate bad patterns for attaining better individuals in the population. Even if the selection methods are unlike but the basic procedure is identical by selecting best individuals for next generation.

2. Genetic Operators - In the inception the first chromosomes that fit better in contrast to other are delineated and remaining population is carried out by using Genetic Operations. The fitness of current chromosomes verified and examined with the bad chromosomes of the previous generation so that we accomplish best results.

a) Mutation – Modifying bits of some chromosome within a population.

b) Crossover – Cross over two offspring to engender new one. Without performing crossover, the identical copy of parent is acquired.

1.2 NEURAL NETWORK

Neural Network advances information in the same way as a human brain do. The network is tranquil with a large amount of interrelated processing neurons, which are functioning in parallel to decipher a particular problem. You can grasp neural network through example. To create composite relationships between input and output data, simply discover neural network. In order to achieve prominent accuracy in prediction of tuberculosis, supervised learning model has been proposed. This learning model has three different layers-input, hidden and the last one is output layer [7]. Back propagation algorithm depends on weight and it predicts the output of the neural network.

The incitement of this analysis is to combine two techniques that are Neural Network and Genetic Algorithm [4]. This study is offered to implement Neural Network with Genetic Algorithm. It predicts the output of function when the network has been trained. A genetic algorithm is used to train a neural network. It does not bother about the way the neural network is connected with each other [5]. In this probe, the weight of neural network is taken as a population which is further utilized by GA to make string and evaluate the fitness function

i. Initialize weights to create one string.
ii. Recognize the network.
iii. Employ a string as a member of the population.
iv. Training is performed until we get predicted result.

Neural network uses parameters such as Cough, Fever, Night Sweats, Chest Pain and Weight Loss, Age, Gender, Haemoglobin, Unwillingness For work, Loss of appetite as input, compute weights and ascribe values at random and generate output.

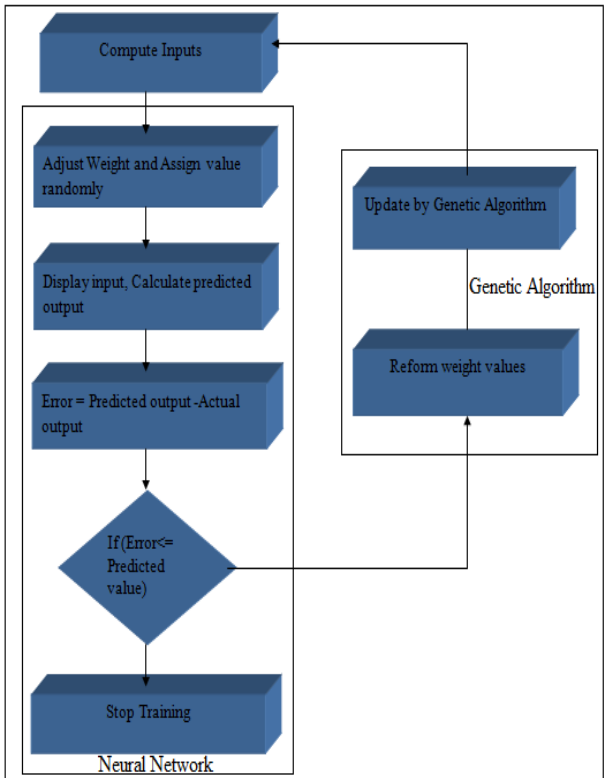


Fig 1. Flowchart of Training Neural Network by Genetic Algorithm

Mean Square Error Function (MSE) is used in a neural network to gauge the difference between predicted and actual output. Every individual of the population takes all weights of the neural network. To figure out individual fitness value, the fitness function is introduced. As a result, the genetic algorithm (GA) gets the better fitness value by the mean of the selection, crossover, and mutation.

If the error is greater than predicted value, Mean Square Error function is engaged by genetic algorithm unless the minimum error is obtained. It will upgrade value by GA continuously until the needed result is found. Further, the Back propagation neural network gets the best possible weights and threshold is given by Genetic algorithm.

The Rest of the paper is arranged as follows. Section II shows related work using Genetic algorithm and Neural Network Approach, Section III shows various steps performed during tuberculosis pre-prediction and finally in section IV about conclusion and future scope of article.

II. LITERATURE REVIEW

Shakshi Garg et. al [1] discussed previous tuberculosis classification algorithm such as Thresholding, color segmentation, Histogram equalization. The aspire of this survey is to examine miscellaneous data mining

DOI 10.17148/IJARCCE.2016.5611

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319 5940
International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 6, June 2016

approaches for tuberculosis detection. To ameliorate the outcome of foregoing methods like SVM, two different techniques are presented such as genetic algorithm and Neural Network. It is congruous where treatment possibilities are less.

D valley et. al [2] proposed the Hybrid system for diagnosis chest diseases using neural network and Genetic Algorithm. The aim of this disease calculator is to insert test report of the user along with their symptoms, to ensure the highest occurring disease among patients. To compare output accuracy of chest diseases such as Asthma, tuberculosis, lung cancer etc., Back propagation algorithm (ANN) and Genetic algorithm were implemented. The genetic algorithm creates new population from the previous one. Back propagation algorithm computes gradient error and adjusts weight values. In this analysis, numerous types of symptoms are taken as input and then converted into values. To gauge the activation function, these values are used.

P. V. Geetha et. al [10] proposed hybrid Genetic-Neuro expert system to classify tuberculosis disease dataset. A genetic algorithm is suitable to determine optimization problem. Artificial Neural Network is trained with Levenberg Marquardt algorithm. In this probe, Genetic-Neuro approach uses nine symptoms of Tuberculosis as input to diagnose TB. Genetic Algorithm and Neural Network together performs better as compared to artificial neural network alone. At the end, GA-NN achieves higher accuracy for tuberculosis classification.

M.k Osman et. al [3] introduced genetic algorithm neural network for diagnosis of mycobacterium tuberculosis in Ziehl-Neelsen stained tissue. The preeminent of this research is to help pathologist for analysis of TB. It supplants manual screening method with a new one; which takes less time as compared to screening method. Z-N stained tissue slides images were captured by the digital camera which is further used for diagnosis of TB. Genetic Algorithm Neural Network approach produces the highest accuracy of 89.64% when compared to Neural Network.

Mumini Olatunji Omisore et. al [13] proposed introduced Genetic- Neuro-Fuzzy Inferential method for diagnosis of tuberculosis. It is an epidemic disease which can inject persons of all age group such as adults, kids etc. If it is not treated on right time, chances of demise increase. This investigation uses record of ten patients and also aids the doctor to analyze tuberculosis in well-timed and economical manner. Results revealed that Genetic-Neuro-Fuzzy Inferential method show enhanced in diagnosis accuracy.

III.PROPOSED WORK

Initially congregate the data which is necessary to affect the tuberculosis. Selection of parameters is the challenging problem. So simply introduce the genetic algorithm and neural network to extract imperative parameters and apply fitness function for each parameters then crossover and training and testing of parameters for tuberculosis detection. Neural network training is done by genetic algorithm. In the end, analyze the output.

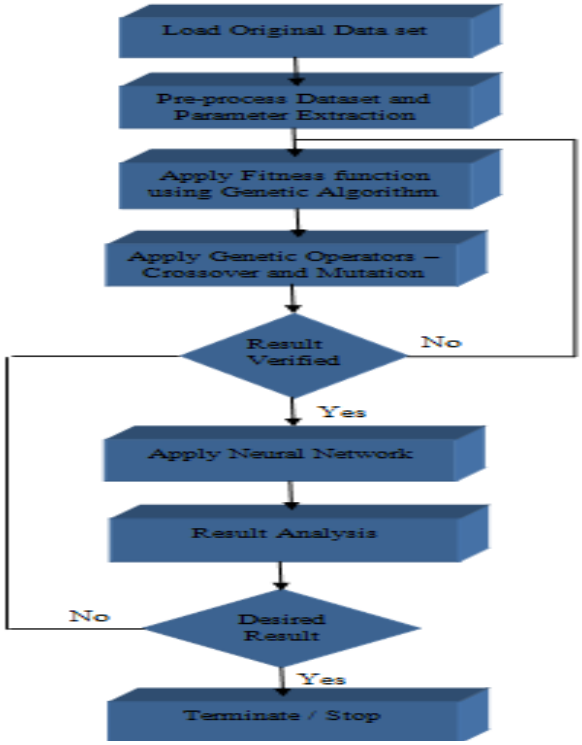


Fig 2. Proposed Algorithm for Tuberculosis Detection

The steps followed in proposed work used for predicting tuberculosis

Step 1: Haul Tuberculosis Disease Dataset.

Step 2: Pre-processing on data and elect 8 imperative parameters by feature selection algorithm.

Step 3: Employ Genetic algorithm to predict tuberculosis. In Genetic Algorithm, three noteworthy features are discussed such as Fitness, Crossover and Mutation.

(i) Fitness

(ii) Crossover

(iii) Mutation

Step 4: Exert neural network by Genetic Algorithm

(i) Levenberg Marquardt algorithm for training

(ii) Carry out Testing

Step 5: Scrutinize the upshot to ensure whether predicted outcome procured or not.

Step 6: If we wangle the desired the backwash then simply cease the process.

Step 7: Validate the result.

IV. PERFORMANCE ANALYSIS

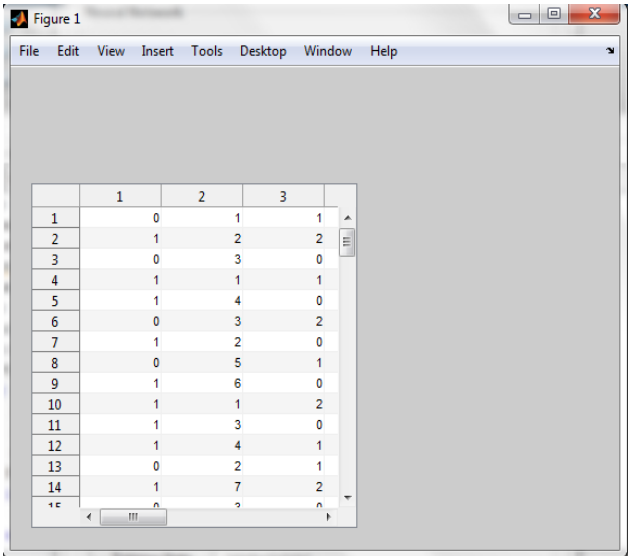
This section shows snapshots of proposed work such as Neural network training window, Training state, Best validation performance plot, Regression plot, Error Histogram and Confusion Matrix.

Figure 3 shows input parameters which are taken for neural network training and testing.

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 6, June 2016



	1	2	3
1	0	1	1
2	1	2	2
3	0	3	0
4	1	1	1
5	1	4	0
6	0	3	2
7	1	2	0
8	0	5	1
9	1	6	0
10	1	1	2
11	1	3	0
12	1	4	1
13	0	2	1
14	1	7	2

Fig. 3 Analysis of Input Parameter

Feed Forward Neural Network model is extensively used for predicting purpose. To diagnose tuberculosis disease, a feed forward back propagation neural network is introduced. In figure 4, Neural network consists of three different layers one is input layer, next is hidden and last one is output layer. Five hidden layer neurons are produced and trained.

The information goes only in forward direction, from the input nodes, passes through hidden nodes and last to the output nodes. In this type of network no loops or cycle exists; hence there is no feed backward concept.

During training phase, hidden neurons would be able to learn patterns in data and provides mapping between input and output nodes.

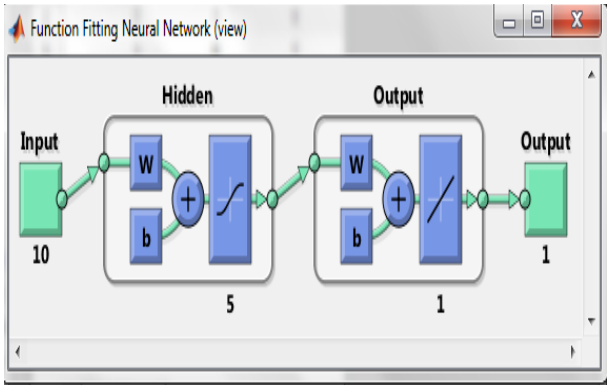


Fig. 3 Function Fitting Neural Network

Each and every hidden node employs transfer function to check acquired input data and passes this processed data to output nodes for additional work.

In figure 5, neural network seize ten inputs and five hidden nodes and generates one output. Neural network training window divides data using divderand function. Training is applied using Levenberg-Marquardt (trainlm) and used with mean square error function.

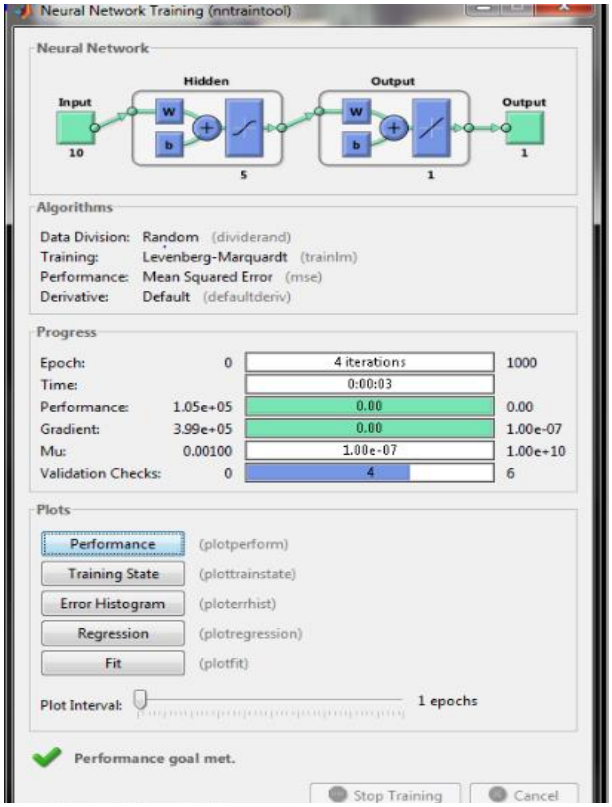


Fig. 5 Neural Network Training Window

Basically in figure 6, more epochs of training represent less error but over-fitting of training data begins when validation data set increases. It means best performance is achieved if we have a minimum validation error. In figure 6, Best validation performance is attained at epoch 4 and represented by dotted line.

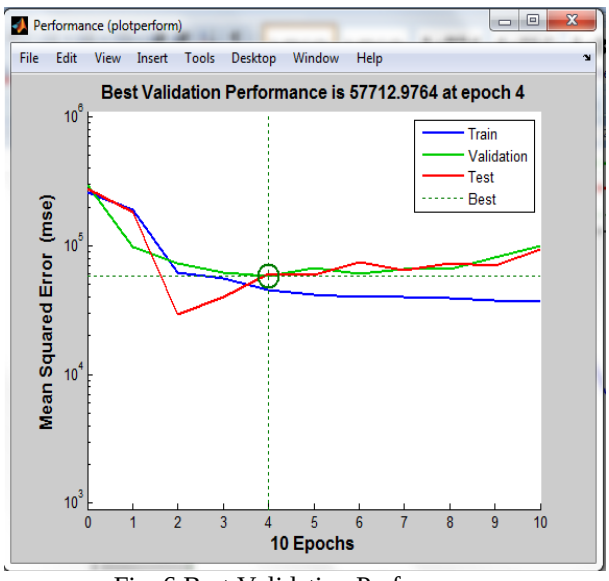


Fig. 6 Best Validation Performance

Figure 7 represents fluctuation in gradient coefficient and number of epochs. Training and Testing of networks will operate better with a minimum value of the gradient coefficient. Gradient value is inversely proportional to a number of epochs. If gradient value increases, the number

DOI 10.17148/IJARCCE.2016.5611

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 6, June 2016

of epochs decreases. This training state shows validation checks, gradient mutation required for training and testing of the neural network.

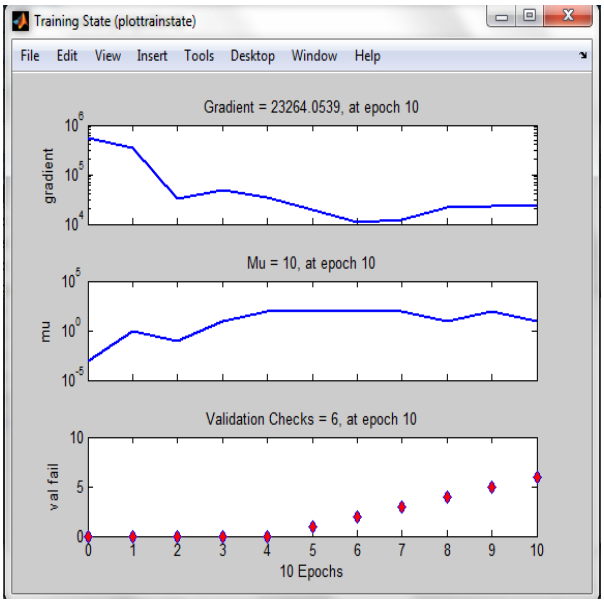


Fig.7. Training State

Figure 8 shows error histogram with 20 Bins. It represents the number of instances with respect to number of errors. This figure indicates zero error. A zero error is only possible if mean square error function generates a minimum error.

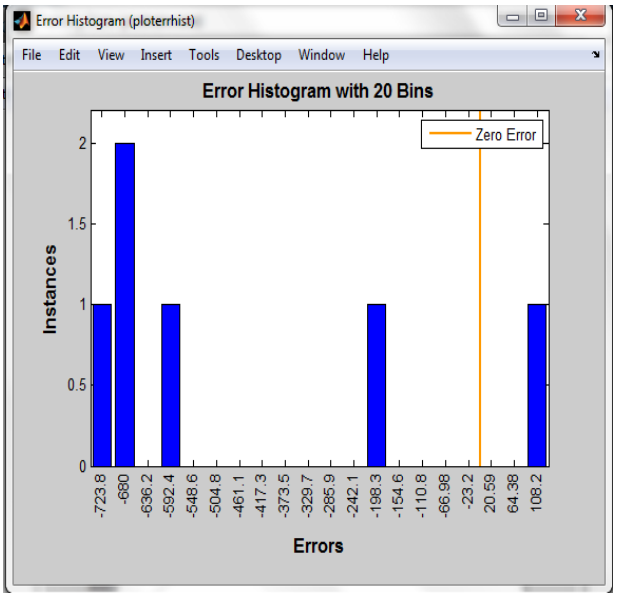


Fig. 8. Error Histogram

To check the network performance, regression plot is introduced. For training sets, it shows output with respect to the target. If network outputs are same as target then it is termed as a perfect fit. In figure 9, fits is fairly acceptable for training sets. Regression plots represents r value = 0.96. If you meliorate more exact results, you can retrain and generate improved network by modifying initial weights.

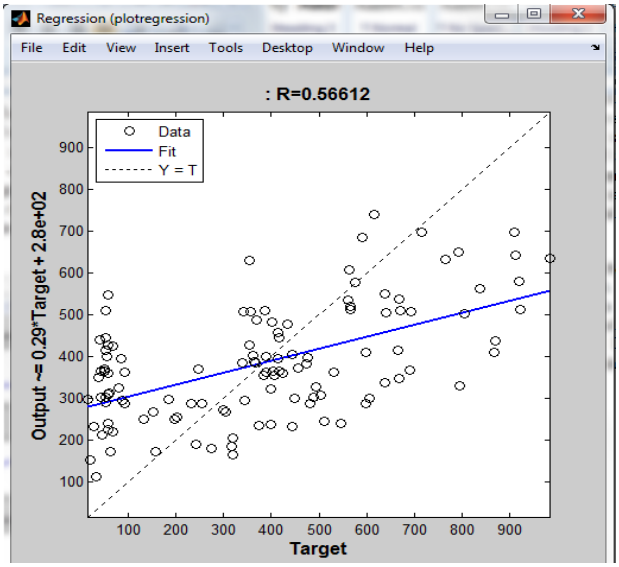


Fig.9. Regression Plot

Figure 10 shows confusion matrix. In this matrix, X-axis represents output class and Y-axis target class. Different square type boxes shows correct and incorrect precision and also shows the highest accuracy. In simple terms, Feed-Forward Back propagation network shows percent of correctly classified in confusion matrix.

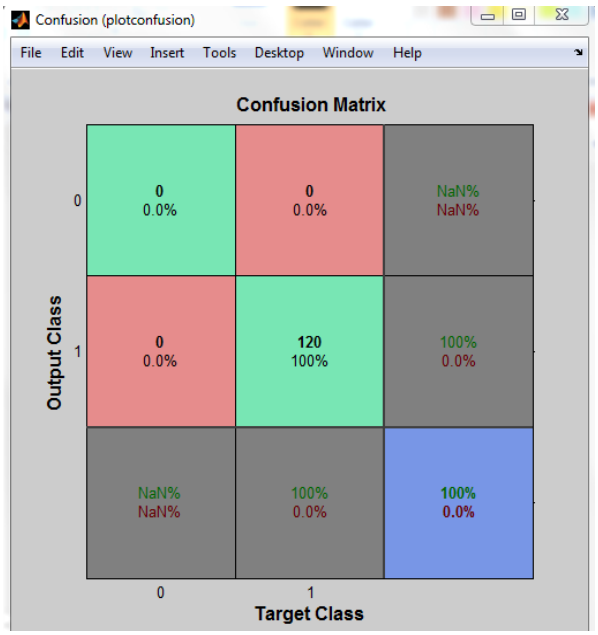


Fig 10. Confusion Matrix

V. CONCLUSION

Soft computing can be applied to detect tuberculosis. For different parameters such as Age, Cough, Chest Pain, Fever, Night Sweats, Chest Pain, Haemoglobin and Weight Loss, Unwillingness for work, Loss of appetite is taken as input parameters. Genetic Algorithm and Neural Network (GANN) performs better with these parameters as compared to other technique. GANN is a very cost effective technique for the prediction of Tuberculosis in a human being.

DOI 10.17148/IJARCCE.2016.5611

VI. FUTURE SCOPE

There are much more different types of Soft Computing Techniques pending like flower pollination, Gravitational Search Algorithm, Artificial fish swarm algorithm. These algorithms can be applied for efficient results for disease prediction.

REFERENCES

[1] Shakshi Garg, Navpreet Rupal," A Review on Tuberculosis Using Data mining Approaches," In 2015 IJEDR, Volume 3, Issue 3, pp.1-4.
[2] D vally and CH V Sarma," Diagnosis Chest Diseases Using Neural Network and Genetic Hybrid Algorithm." In 2015, International Journal of Engineering Research and Applications, Volume 5, Issue 1(Part 2), pp.20-26.
[3] M. K. Osman, F. Ahmad, Z. Saad, M. Y.Mashor and H. Jaafar," A Genetic Algorithm-Neural Network Approach for Mycobacterium Tuberculosis Detection in Ziehl-Neelsen Stained Tissue Slide Images," In 2015 tenth International Conference on Intelligent Systems Design and Applications, pp. 1229-1234.
[4] Md. Mijanur Rahman and Tania Akter Setu," An Implementation for Combining Neural Networks and Genetic Algorithms," In 2015 International Journal Computer Science and Technology Vol 6, Issue 3, pp. 218-222.
[5] Ranno Agarwal," Genetic Algorithm in Data Mining," In 2015 IJARCSSE, Vol 5, Issue 9, pp. 631-634.
[6] Monika Gandhi and Shailendra Narayan Singh, "Prediction in Heart disease using techniques of data mining," In 2015 International conference on Futuristic Trend in Computational Analysis and knowledge Management, pp.520-525.
[7] S.Vijayarani and S.Dhayanand, "Liver disease prediction using SVM and Naives Bayes Algorithm," In 2015 International Journal of Science Engineering and Technology Research,Vol-4,Issue 4,pp.816-820.
[8] T. Kamalavalli, S. Vinothini and R. Kuppuchamy," A Data Mining Technique for Breast Cancer Diagnosis," In 2015 IJSRD, Vol 3, Issue 02, pp. 199-202.
[9] Tanmay Tamhane, Mateen Shaikh, Sanjaykumar Boga, Mrunal Tanwar and A.E. Patil," Heart Disease and Diabetes Prediction using Data Mining," In 2015 International Journal of Scientific Research & Development,Vol-3 , Issue-1, pp. 1052-1054.
[10] P. V. Geetha, R. A. Lukshmi and P. Venkatesan," Tuberculosis Disease Classification using Genetic-neuro Expert System," In 2014 Indian Journal of Science and Technology, Volume 7 Issue 4, pp. 421–425.
[11] Rusdah and Edi Winarko," Review on Data Mining Methods for Tuberculosis Diagnosis," In 2013 ISICO.
[12] K. Rama Lakshmi and S.Prem Kumar," Utilization of Data Mining Techniques for Prediction and Diagnosis of Major Life Threatening Diseases Survivability-Review," In 2013 International Journal of Scientific & Engineering Research, Vol 4,Issue 6,pp. 923-932.
[13] Mumini Olatunji Omisore, Oluwarotimi Williams Samuel and Edafe John Atajeromavwo," A Genetic-Neuro-Fuzzy inferential model for diagnosis of tuberculosis," Applied Computing and Informatics, In Press.

DOI 10.17148/IJARCCE.2016.5611