RESEARCHERID
THOMSON REUTERS

[Shrivastava* *et al.*, 6(7): July, 2017]
IC™ Value: 3.00

ISSN: 2277-9655
Impact Factor: 4.116
CODEN: IJESS7

IJESRT

INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY

DIAGNOSIS OF HEART DISEASE USING GENETICALLY OPTIMIZED NEURAL NETWORK

Shiva Shrivastava[*1] & Neeraj Mehta[2]

DOI: 10.5281/zenodo.823060

ABSTRACT

Cardiovascular diseases are a major public health problem and are expected to continue to be in the future due mainly to population aging; In Spain, represent the leading cause of death and hospitalization. The three most important cardiovascular problems, ischemic heart disease, cerebrovascular disease and heart failure are based on the large epidemiological studies.This paper proposes Genetically optimized Neural Network technique for heart disease prediction. Performance of proposed approach is evaluated using confusion matrix plot.

KEYWORDS: Back Propagation, Cardiovascular disease Confusion Matrix, Genetic Algorithm, Neural Network.

I. INTRODUCTION

Cardiovascular diseases are a major public health problem. Such a claim corresponds to a current reality, however, it is anticipated that cardiovascular pathologies will continue to be an important health problem in the future, mainly due to the increase in their prevalence in developing countries and population aging [1].

Around the world, 16.7 million annual deaths can be attributed to cardiovascular disease. More specifically, in high-income countries, heart disease and cerebrovascular disease represent the first and second leading cause of death among men and women [2]. In Europe alone, in 2008, more than 4.3 million deaths were attributed to cardiovascular diseases, which represented 48% of the total number of deaths on the continent [3]

However, despite the serious problem that cardiovascular diseases currently pose, knowledge of their main modifiable risk factors makes prevention possible. The three most important modifiable cardiovascular risk factors are smoking, hypertension and hypercholesterolemia, followed by diabetes, overweight / obesity, sedentary lifestyle and alcohol abuse [4].

The results research in health allows to evaluate the quality and effectiveness of health care, determined by obtaining pre-established final results. The measurement of such results is possible through well-conceptualized indicators. Avoidable Hospitalizations for Ambulatory Care-Sensitive Conditions (ACSCs) is a sanitary indicator that, through the quantification of hospitalizations caused by a specific group of pathologies (including cardiovascular), aims to measure the primary care capacity, based on the reason that, the increase of preventive measures and the improvement of the ambulatory treatments in this level of attention, should correspond with a reduction of those hospitalizations. The objective of this study is to analyze the effect that the preventive programs of the cardiovascular diseases applied in primary care have on the avoidable hospitalization specific of these diseases.

II. CARDIOVASCULAR PREVENTION

Cardiovascular disease (CVD) is defined as a group of conditions that affect the heart and blood vessels. Despite the decreasing trend in developed countries in the last three decades, CVD as a whole is the main cause of mortality and hospitalization in the Spanish population [5]. The three basic cardiovascular problems, ischemic heart disease, cerebrovascular disease and heart failure are based on the knowledge of the large epidemiological studies developed since the middle of the last century. The Framingham Heart Study, has become the most classic cohort study by antonomasia, establishing for more than six decades the essential role played by risk factors in the development of CVD. Primary cardiovascular prevention is based on the identification and control of cardiovascular risk factors in the healthy population, thus attempting to prevent the onset of the disease [6].

CVDs show differences between men and women. While women are the number one cause of death in women, CVD is the second cause of death in men. On the other hand, the trend in hospital morbidity rates of CVD in recent years has been steadily increasing, both in men and women. In this sense, cardiovascular diseases are the leading cause of hospitalization in the population. In the coming years, there is an increase in the number of hospitalizations for these diseases, as a result of the technological development that will allow patients to offer new diagnostic and therapeutic instruments, the greater survival of patients with these health problems and the aging of the Spanish population [7].

In 2009, hospitalizations for CVD recorded the highest number of hospital discharges (12.8 per 100 high). CVD were the second cause of hospitalization among women (10.6%), only surpassed by pregnancy, delivery and puerperium. In men, it was the first cause of hospitalization, causing 15.3% of the total number of care incidences in that year [8].

III. PROPOSED METHODOLOGY

The problem with risk factors related to heart disease is that there are many risk factors involved like age, usage of cigarette, blood cholesterol, person's fitness, blood pressure, stress and etc. and understanding and categorizing each one according to its importance is a difficult task. Also a heart disease is often detected when a patient reaches advanced stage of the disease. Hence the risk factors are analyzed from various sources [9]-[10]. The dataset was composed of 12 important risk factors which were sex, age, family history blood pressure, Smoking Habit, alcohol consumption, physical inactivity, diabetes, blood cholesterol, poor diet, obesity .The system indicated whether the patient had risk of heart disease or not. The data for 50 people was collected from surveys done by the American Heart Association [10]. Most of the heart disease patients had many similarities in the risk factors [11]. Table 4.1 shows the identified important risk factors and the corresponding values and their encoded values in brackets, which were used as input to the system.

Table 1: Risk factors values and their encodings [12]

S. No.	Risk Factors	Values
1	Sex	Male (1), Female (0)
2	Age (years)	20-34 (-2), 35-50 (-1), 51-60 (0), 61-79 (1) , >79 (2)
3	Blood Cholesterol	Below 200 mg/dL - Low (-1) 200-239 mg/dL - Normal (0) 240 mg/dL and above - High (1)
4	Blood Pressure	Below 120 mm Hg- Low (-1) 120 to 139 mm Hg- Normal (0) Above 139 mm Hg- High (-1)
5	Hereditary	Family Member diagnosed with HD -Yes (1) Otherwise – No (0)
6	Smoking	Yes (1) or No (0)
7	Alcohol Intake	Yes (1) or No (0)
8	Physical Activity	Low (-1) , Normal (0) or High (-1)
9	Diabetes	Yes (1) or No (0)
10	Diet	Poor (-1), Normal (0) or Good (1)
11	Obesity	Yes (1) or No (0)
12	Stress	Yes (1) or No (0)
Output	Heart Disease	Yes (1) or No (0)

Data analysis has been carried out in order to transform data into useful form, for this the values were encoded mostly between a range [-1, 1]. Data analysis also removed the inconsistency and anomalies in the data. This was needed. Data analysis was needed for correct data pre-processing. The removal of missing and incorrect inputs will help the neural network to generalize well.

ISSN: 2277-9655
Impact Factor: 4.116
CODEN: IJESS7

In this paper, geneticallyoptimized Neural Network approach is used to determine optimum number of clusters in analyzed data. These methods are described below.

Diagnosis using Genetically Optimized Neural Network

In previous phase, Neural Network is trained using back-propagation algorithm to find weight and bias values. The proposed GA-NN approach uses Genetic Algorithm to find weight and bias values. In this proposed NN, the value of weight and bias are random and to correct these values a fitness function is employed for Genetic Algorithm.

Fitness Function: The fitness function is a function of weight and bias with the objective of minimizing the mean square error between the predicted and target classes of the training data.

$$\min F(w,v) = \sum_{t=1}^{q}[c_t - (wx_t + v)]^2 \qquad (1)$$

Where, x_t is input and c_t is target output.
Fitness function in equation (1) is minimized using Genetic Algorithm to optimized weight and bias values.

Genetic algorithms (GA) are a fairly wealthy family and very interesting stochastic optimization algorithms based on the mechanics of natural selection and genetics. The choice of GA among other methods is justified based on the following four properties [13]:

- The GA use encryption settings and not the settings themselves.
- The GA are working on a population of points, rather than a single point.
- The GA use only the values of the studied function, not its derivative or other auxiliary knowledge.
- The GA use probabilistic transition rules, not deterministic.

In addition, the GA using two major strategies to find a solution or set of solutions. These strategies are: exploration and exploitation. They allow to find the global maximum (solve the problem) because they are complementary [14]. If the exploration is investigating all solutions of the search space, the operational phase in turn uses the found knowledge to previously visited solutions to help find better solutions. The combination of these two strategies can be quite effective but the difficulty is to know where the best solution is.

The GA operate with a population comprising a set of individuals called chromosomes. Each chromosome is composed of a set of genes. Each individual is assigned a value calculated by a function called adaptive or fitness. In practice, from a population of chromosomes is generated in a random manner during initialization. To set the size of the population, Marczyk, A. [14] reported that this size varies from one problem to another. In each cycle of genetic operations, a new population called generation is created from the chromosomes of the current population. Why some chromosomes called 'parents' are selected to develop the genetic operations. The genes of these parents are mixed and recombined to produce other chromosomes called 'children' forming the new generation. The steps of the GA are repeated in cycles t, the judgment of the algorithm is fixed according to a stopping criterion. There can be several stopping criteria:

- The number of generation originally set has been reached.
- The value of the adaptation function has reached a set value a priori.
- The absence of changes in the value of the fitness function of individuals in a population to another.
- The chromosomes have reached a certain degree of homogeneity.

ISSN: 2277-9655
Impact Factor: 4.116
CODEN: IJESS7

Figure 1 shows the steps of a simple GA:

```
Simple Genetic Algorithm
{ t = 0;
        Initialization (P (t));
        Evaluation (P (t));
        While not end to t =
        t+1;
            P'= select (P (t));
          Recombining (P'(t);
            Mutation (P'(t));
          Evaluation (P '(t));
             P = Survival (P, P'(t));
}
```

Figure 1: Steps of a simple GA

Figure 2 shows the evolutionary cycle for Genetic algorithm

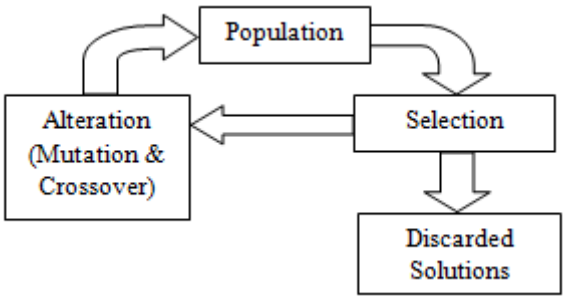

Figure 2: Genetic algorithm evolutionary cycle [14]

The data for risk factors related to heart diseases collected from 50 people is provided in Table 2.

Table 2: Patient's case study data in encoded form

No	Sex	Age	Blood Cholesterol	Blood Pressure	Hereditary	Smoking	Alcohol Intake	Physical Activity	Diabetes	Diet	Obesity	Stress	Heart Disease
1	F	35	High	Normal	No	No	Yes	Low	Yes	Poor	Yes	Yes	Yes
2	M	70	Low	Low	No	No	Yes	High	Yes	Normal	No	No	No
3	F	60	High	High	No	No	No	Normal	Yes	Poor	Yes	Yes	Yes
4	F	36	Low	Normal	No	No	No	Normal	No	Good	No	No	No
5	M	30	Low	Normal	No	No	Yes	High	No	Normal	No	No	No
6	F	39	Low	Normal	Yes	No	Yes	High	Yes	Normal	No	Yes	No

ISSN: 2277-9655
Impact Factor: 4.116
CODEN: IJESS7

7	F	41	High	Normal	No	No	No	Low	No	Poor	Yes	No	No
8	M	70	High	Normal	No	No	Yes	Low	No	Poor	Yes	No	Yes
9	M	65	Normal	High	Yes	Yes	Yes	Normal	Yes	Poor	Yes	No	Yes
10	M	30	Normal	High	No	Yes	No	Normal	No	Good	No	Yes	No
11	F	31	Low	Normal	No	No	No	High	No	Normal	No	No	No
12	F	29	Low	Normal	No	No	Yes	High	No	Good	No	No	No
13	M	30	Low	Normal	No	No	Yes	Normal	No	Normal	No	No	No
14	F	45	Normal	High	Yes	Yes	No	Normal	Yes	Normal	Yes	Yes	No
15	M	25	High	Normal	Yes	Yes	Yes	Low	Yes	Normal	No	No	Yes
16	F	37	Normal	Normal	No	No	No	Normal	Yes	Poor	No	Yes	No
17	F	37	Normal	High	No	Yes	Yes	High	No	Poor	No	Yes	No
18	M	53	High	Low	No	Yes	No	Normal	Yes	Normal	No	Yes	No
19	M	57	High	Normal	No	Yes	No	Low	No	Poor	Yes	Yes	Yes
20	M	52	High	Low	No	No	No	Normal	Yes	Poor	Yes	No	No
21	M	48	Normal	Normal	Yes	Yes	Yes	Normal	No	Normal	No	No	Yes
22	M	62	High	High	No	Yes	Yes	Normal	Yes	Normal	No	No	Yes
23	M	56	Normal	High	No	Yes	Yes	Low	No	Poor	Yes	Yes	Yes
24	F	27	Low	Normal	No	No	No	High	No	Good	No	No	No
25	M	33	Normal	Normal	No	No	No	Normal	Yes	Good	No	No	No
26	F	33	Normal	Normal	No	No	Yes	Low	Yes	Poor	No	Yes	No
27	M	37	High	Normal	No	No	Yes	Normal	No	Normal	No	Yes	No
28	M	43	Normal	High	No	No	No	Normal	Yes	Poor	Yes	Yes	Yes
29	M	46	Low	Normal	No	No	No	Normal	Yes	Poor	Yes	Yes	No

ISSN: 2277-9655
Impact Factor: 4.116
CODEN: IJESS7

30	F	36	Low	Normal	No	No	No	Normal	No	Normal	No	No	No
31	F	29	Low	Normal	No	No	No	Normal	No	Good	No	No	No
32	F	47	Normal	Normal	No	No	Yes	High	Yes	Normal	No	Yes	No
33	M	58	High	High	No	Yes	Yes	Normal	Yes	Normal	No	Yes	Yes
34	M	44	High	Normal	Yes	Yes	Yes	Normal	No	Normal	Yes	Yes	Yes
35	F	36	Normal	High	No	No	No	Normal	No	Good	Yes	No	Yes
36	M	42	Low	Normal	Yes	No	Yes	Low	No	Poor	No	Yes	No
37	F	25	Low	Normal	No	No	No	High	No	Poor	No	No	No
38	F	28	Low	Normal	No	No	Yes	High	No	Normal	No	No	No
39	F	26	Low	Normal	Yes	No	No	Normal	No	Normal	Yes	No	Yes
40	M	28	Low	Normal	No	No	No	Normal	No	Poor	No	No	No
41	F	45	High	Normal	No	No	Yes	Low	Yes	Poor	Yes	Yes	Yes
42	M	63	Low	Low	No	No	Yes	High	Yes	Good	No	No	No
43	F	55	High	High	No	No	No	Normal	Yes	Normal	Yes	Yes	Yes
44	F	44	Low	Normal	No	No	No	Normal	No	Normal	No	No	No
45	M	35	Low	Normal	No	No	Yes	High	No	Normal	No	No	No
46	F	42	Normal	Normal	No	No	Yes	High	Yes	Good	No	No	No
47	F	43	Normal	Normal	No	No	No	Low	No	Poor	Yes	No	No
48	M	65	Normal	Normal	No	No	Yes	Low	No	Normal	Yes	Yes	Yes
49	M	74	Normal	High	No	Yes	Yes	Normal	Yes	Normal	Yes	Yes	Yes
50	M	36	Normal	High	No	Yes	No	Normal	No	Poor	No	No	No

ISSN: 2277-9655
Impact Factor: 4.116
CODEN: IJESS7

IV. SIMULATION AND RESULTS

The performance of proposed technique has been studied by means of MATLAB simulation.

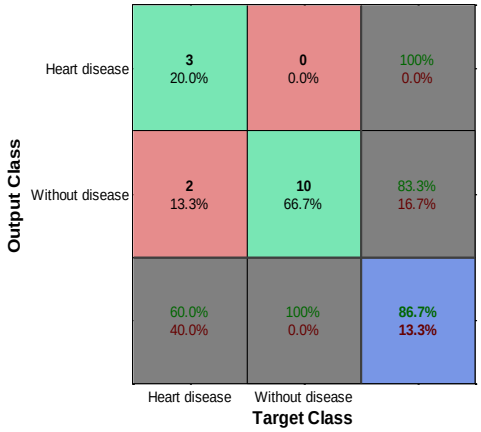

Figure 3: Confusion matrix for NN method

The row and column are the labels for detection and no detection of heart disease from database. There are 2 sets of classes and each class having different set of detection. Total 15 samples of patients are taken out of which 3 patients are classified correctly and none of the samples were misclassified but in normal category 10 samples are classified correctly out of 15 samples and 2 samples are misclassified. The confusion plot indicates the accuracy i.e. 86.7% for this approach.

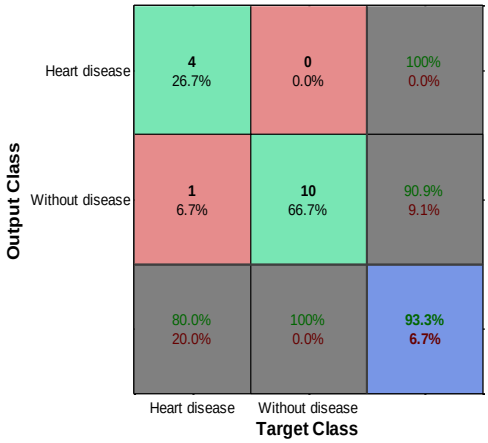

Figure 4: Confusion matrix for GA-NN method

The row and column are the labels for detection and no detection of heart disease from database. There are 2 sets of classes and each class having different set of detection. Total 15 samples of patients are taken out of which 4 patients are classified correctly and none of the samples were misclassified but in normal category 10 samples are classified correctly out of 15 samples and 1 sample is misclassified. The confusion plot indicates the accuracy i.e. 93.3% for this approach.

- Population Size: 30
- Iteration: 500

V. CONCLUSION

In this paper, Genetic Algorithm uses the phenomena of mutation and crossover over various generations. The weights which are used for back-propagation can be optimized first and then given as input to our network to give much better results and it was found that it is effective to predict the risk of heart disease when the person provide the required attributes value. On observing the confusion matrix, it was found that the GA-NN based approach outperforms the NN based approach with the accuracy of 93.3%.

ISSN: 2277-9655
Impact Factor: 4.116
CODEN: IJESS7

VI. REFERENCES

[1] Jones, D.S. and Greene, J.A., 2013. The decline and rise of coronary heart disease: understanding public health catastrophism. *American journal of public health*, *103*(7), pp.1207-1218.
[2] Mathers, C.D. and Loncar, D., 2006. Projections of global mortality and burden of disease from 2002 to 2030. *Plos med*, *3*(11), p.e442.
[3] World Health Organization, 2011. *Global status report on noncommunicable diseases 2010*. Geneva: World Health Organization.
[4] Buttar, H.S., Li, T. and Ravi, N., 2005. Prevention of cardiovascular diseases: Role of exercise, dietary interventions, obesity and smoking cessation. *Experimental & Clinical Cardiology*, *10*(4), p.229.
[5] Jenkins, C.D., 1988. Epidemiology of cardiovascular diseases. *Journal of consulting and clinical Psychology*, *56*(3), p.324.
[6] World Health Organization, 2007. *Prevention of cardiovascular disease*. World Health Organization.
[7] Cowie, M.R., Anker, S.D., Cleland, J.G., Felker, G.M., Filippatos, G., Jaarsma, T., Jourdain, P., Knight, E., Massie, B., Ponikowski, P. and López-Sendón, J., 2014. Improving care for patients with acute heart failure: before, during and after hospitalization. *ESC Heart Failure*, *1*(2), pp.110-145.
[8] Ajay, V.S. and Prabhakaran, D., 2010. Coronary heart disease in Indians: Implications of the INTERHEART study. *The Indian journal of medical research*, *132*(5), p.561.
[9] Centre for Disease Control and Prevention, Online available at: http://www.cdc.gov/heartdisease/risk_factors.htm
[10] American Heart Association, Online available at: http://www.heart.org/HEARTORG/Conditions
[11] D. Isern, D. Sanchez, and A. Moreno, "Agents Applied in Health Care: A Review", International Journal of Medical Informatics, Vol. 79, Issue 3, pp. 146-166, 2010.
[12] Syed Umar Amin, Kavita Agarwal, Dr. Rizwan Beg, "Genetic Neural Network Based Data Mining in Prediction of Heart Disease Using Risk Factors", Proceedings of IEEE Conference on Information and Communication Technologies (ICT), pp. 1227–1231, April 2013.
[13] Forrest, S., 1993. Genetic algorithms- Principles of natural selection applied to computation. *Science*, *261*(5123), pp.872-878.
[14] Marczyk, A., 2004. Genetic algorithms and evolutionary computation. *The Talk Origins Archive:* http://www.talkorigins/faqs/genalg/genalg.html

CITE AN ARTICLE

Shrivastava, S., & Mehta, N. (2017). DIAGNOSIS OF HEART DISEASE USING GENETICALLY OPTIMIZED NEURAL NETWORK. *INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY*, *6*(7), 142-149. doi:10.5281/zenodo.823060