

# Genetically Optimized Neural Network for Heart Disease Classification

Snehal Patil<sup>1</sup> Manish Patil<sup>2</sup>

<sup>1&2</sup> Department of Electronics and Telecommunication Engineering, Gangamai College of Engineering, Nagaon, Dhule, MS, India

\*\*\*

**Abstract** – This study seeks to implement and analyze a methodology for Heart Disease (Cardiovascular disease) detection using machine learning, based on quantitative procedures, to assist in the diagnosis of individuals with heart disease, through the investigation and selection of variables that have a greater degree of importance in determining this event. The diagnosis of heart disease is performed using a model of Neural Networks after a sequence of pre-treatment procedures and selection of variables. Further neural network is optimized with genetic algorithm to get better accuracy. Confusion matrix is plotted for accuracy.

**Keywords** – Back Propagation; Cardiovascular disease; Confusion Matrix; Genetic Algorithm; Neural Network.

## 1. INTRODUCTION

A sedentary lifestyle, high cholesterol and stress are some of the heart's worst enemies. They fit within the so-called modifiable factors, that is, they are repairable from changes in the population's habits. Some people are more likely to develop heart disease, especially given their age and family history. But that doesn't mean younger people shouldn't be worried.

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 [1].

Artificial Intelligence (AI) applied to medicine is a resource important in solving health care problems and comes becoming an essential component of medical informatics. Growing increase in challenges potentiated by longer life expectancy, inadequate lifestyle, inequality in access to health services, increased expenses and shortage of professionals of health, demand technological updates, with robust forecast values and alternative scenarios that can map future health problems and options to modify these trajectories [2]. With the growing demand for computational methods that can contribute to diagnoses and optimize the time of health professionals, several researches related to the application of Machine Learning (ML) to health have

been developed [3]. ML is a sub-area of research in AI, which aims to develop computational techniques on learning, as well as the construction of systems capable of acquiring knowledge automatically [4]. In addition, he has the ability to learn from large volumes of data and generate useful hypotheses [5], which can help the health professional in the diagnosis, prevention and search for more effective treatments. The ML model has all the necessary information about a problem, such as, what data constitute the analysis and also what it is expected to produce as knowledge [6]. In addition, ML can also potentially predict the likelihood that future subjects will have Specific diseases, given early screening from the data routine physical and laboratory examinations.

In the context of the present research, learning algorithms of machine can contribute to the prediction of such diseases according to with the characteristics of the individual. However, you must select among various machine learning algorithms the one that best fits the context of the problem. In view of this, the main the objective of the research was the evaluation and comparison of performance of different ML techniques applied to the prediction of cardio metabolic diseases. The main motivation of this work is the study of data dimensionality reduction techniques in order to 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.

## 2. LITERATURE

In medicine, there are implementations of predictive algorithms with techniques of Machine Learning and Data Mining in different fields. There are investigations and even several implementations to detect cancerous tumours, classify the cancer of accompaniment to different categories of diagnoses, even to detect the heart diseases [7]. There are different algorithms that are used for predictive systems and each one throws results with different degrees of precision depending on the data used and on them expected results.

There are 3 types of learning algorithms within Machine Learning:

Supervised, unsupervised and by semi supervised. Depending on the needs from the problem to be solved, the environment and what will develop and the factors that will affect decision-making [8], each of these types of learning in the offers a better approach.

The objective is to develop a predictive system to determine patients who can be prone to suffering heart diseases, through the use of the algorithms supervised, Naive Bayesian [9] and Semi-Naive Bayesian and compare these results with the existing Machine learning techniques [10]. The idea is to train the system with real patient data from which medical data are taken (diabetes, etc.) and personal data (age, sex) to be able to determine there, according to certain characteristics, which people can be prone to suffer some type of illness from heart diseases.

### 3. PROPOSED METHODOLOGY

The problem with risk factors related to heart disease is that it includes many risk factors such as age, cigarette use, blood cholesterol, fitness, blood pressure, stress and so on and it is important as well as difficult task to understand the importance of each and classify it. Also, heart disease is often diagnosed when the patient has reached an advanced stage of the disease. Therefore, risk factors can be analyzed from different 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.

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

#### A. 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.

#### B. 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.

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

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.

#### C. Neural Network

Artificial Neural Network (or simply "Neural Network") is a distributed model composed of units (called

"neurons" in the literature) made up of non-linear functions (typically sigmoid and hyperbolic tangents). The combination of these units, through parameters estimated from the data, is what gives this model the ability to infer non-linear relationships of arbitrary complexity. In the way used in this study, these units are arranged in layers, including a hidden layer, which is not directly connected to the model output. These connections between units, or neurons, are called weights (originally the terminology was "synaptic weights"). These weights are the model parameters that are adjusted by an iterative algorithm through the data. Once the weights are adjusted, the network has the ability to represent the relationship between the input data and the output variable, in this case the diagnosis of disease. The ability to learn from "examples" or data (in-sample) and to generalize (out-of-sample) information generated in complex non-linear environments is undoubtedly the great advantage of Neural Networks.

The relevant and non-redundant variables are used as input to the Neural Network, and, after the training process, the diagnosis of heart disease is output from the Network. Thus, in the training phase, the output of the network assumes values 0 or 1 for each of two possibilities, non-sick or sick, respectively. A sigmoid activation function is used in the output unit so that the Mains output always varies between 0 and 1, since this function saturates these values. In the test phase, and later in the use of the model, values close to 1 indicate a high possibility of the individual being sick and close to 0 indicate a small possibility of being sick. The cut-off point adopted to differentiate high from low possibility was 0.5 and thus only two types of output were considered, sick or non-sick, for values above 0.5 or values below 0.5, respectively.

The relevant and non-redundant variables are used as input to the Neural Network, and, after the training process, the diagnosis of heart disease is output from the Network. Thus, in the training phase, the output of the network assumes values 0 or 1 for each of two possibilities, non-sick or sick, respectively. A sigmoid activation function is used in the output unit so that the Mains output always varies between 0 and 1, since this function saturates these values. In the test phase, and later in the use of the model, values close to 1 indicate a high possibility of the individual being sick and close to 0 indicate a small possibility of being sick. The cut-off point adopted to differentiate high from low possibility was 0.5 and thus only two types of output were considered, sick or non-sick, for values above 0.5 or values below 0.5, respectively.

#### D. Optimization of the Map Weights Set

In this approach, the application of the genetic algorithm will be done using one of the fitness criteria

discussed here. From the formal model defined here for Kohonen networks, we set: the set  $N$  of map positions, a set of input vectors  $X$ , the topology  $T$  of the map and the FA activation and Fw weight adjustment functions, establishing low learning rates for this and initial and final neighbourhood regions. We optimize the set of initial weights of the network  $W(0)$ ; resulting in a network with prior training. The data for risk factors related to heart diseases collected from 50 peoples with risk factor provided in Table 1.

#### E. Genetic Algorithms

Genetic algorithms (Goldberg, 1989), here also called GA, are search methods based on the principles of natural evolution and genetics. In this method, a population of possible solutions to the problem in question evolves under the application of probabilistic operators idealized from biological processes. Thus, there is a tendency for elements of the population become well and a better approximation of the solution as the algorithm develops.

The individuals of the population are represented in a coded way, being common the binary representation, expressed by a string of bits called chromosome. The positions in the chain are called gene and these normally assume the binary values 0 and 1, called alleles. GAs are iterative algorithms, with each iteration called a generation. In most applications the initial population is generated randomly.

GAs assigns to each member of the population a non-negative value of an objective function to be optimized called "fitness" or suitability. This value indicates the quality of the individual in the population, being a measure of its adaptability to the environment. Individuals stronger, with better fitness, will have a better chance of surviving and passing on to the next generation. GA simple is composed of 3 basic operators: selection, crossover or "crossover" and mutation. At selection each individual "i" of the population is selected or not according to a selection probability  $p_i = \frac{f_i}{\sum f_i}$  where  $f_i$  is your "fitness". The selected individuals are grouped into pairs and subjected to the crossover operator. It occurs according to a fixed probability  $P_x$ , simulating sexual reproduction, where chromosome fragments (sub-bit strings) are exchanged.

The mutation occurs according to a probability  $P_m$  and consists of selecting a position in the chain chromosome and changes its binary value. The mutation introduces new genetic material into the population and allows the exploration of new regions in the parameter space. After applying these operators a new population of possible solutions is obtained. The process continues until some stopping criterion. GAs only need information about the value of an objective function, not requiring derivatives or any other kind of knowledge. GAs can be applied to

problems of global optimization, in multidimensional spaces and with objective functions strongly not linear.

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. 90% for this approach.

Optimization of Weight Adjustment Parameters and Map Topology

In this approach, the application of the genetic algorithm will be done using one of the fitness criteria here discussed. From the formal model defined for Kohonen networks, we fix the following set N from map positions, a set of X input vectors; the random set of initial weights W(0); the network training time Tf; and the FA activation and Fw weight adjustment functions, described in the formal specification of the model. The parameters that will be optimized are: the radius of initial and final neighborhood,  $\sigma_i$  and  $\sigma_f$ ; the initial and final learning rate,  $\alpha_i$  and  $\alpha_f$  and the topology T of the map. In this study, the line, grid and cube topologies are used, respectively, forming uni, bi and three-dimensional maps.

```

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));
}
    
```

Fig -1: Genetic algorithm evolutionary cycle

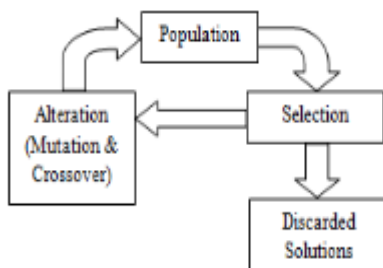


Fig -2: Genetic algorithm evolutionary cycle

4. SIMULATION AND RESULTS

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

	Heart Disease	W/o Heart Disease	
Heart Disease	3 30%	0 0%	100 0%
W/o Heart Disease	1 10%	6 60%	85.7% 14.3%
	75% 25%	100% 0%	90% 10%
	Heart Disease	W/o Heart Disease	

Target Class

Fig- 3: Confusion matrix for NN method

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.

	Heart Disease	W/o Heart Disease	
Heart Disease	3 30%	0 0%	100 0%
W/o Heart Disease	1 10%	6 60%	85.7% 14.3%
	75% 25%	100% 0%	90% 10%
	Heart Disease	W/o Heart Disease	

Target Class

Fig- 4: Confusion matrix for GA-NN method

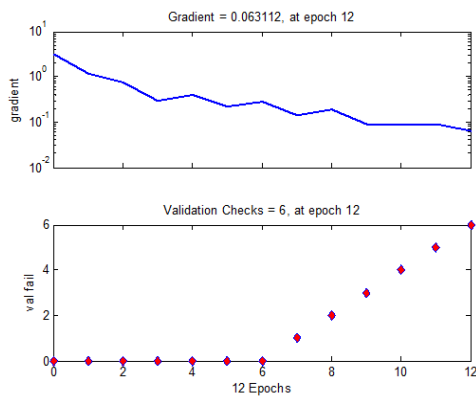


Fig-5: Training model of neural network

## 5. CONCLUSIONS

Adjusting the weight of the original neural network model requires high computing costs because we adjust all the weights of the map for each vector of the training phase. So, the approach of adjusting the weight adjustment parameters for each population generated by the genetic algorithm, according to the desired fitness function, is useful to identify which set of weight adjustment parameters can be used in the conventional model. It can be seen that the cultural algorithm gives similar results to the GA-NN approach i.e. 93.3 but at the lower iterations. Therefore, the Cultural algorithm based approach is best among all three approaches.

## REFERENCE

- [1] Safdar, S., Zafar, S., Zafar, N., & Khan, N. F. (2018). Machine learning based decision support systems (DSS) for heart disease diagnosis: a review. *Artificial Intelligence Review*, 50(4), 597-623.
- [2] Al'Aref, S. J., Anchouche, K., Singh, G., Slomka, P. J., Kolli, K. K., Kumar, A., ... & Min, J. K. (2019). Clinical applications of machine learning in cardiovascular disease and its relevance to cardiac imaging. *European heart journal*, 40(24), 1975-1986.
- [3] Otoom, A. F., Abdallah, E. E., Kilani, Y., Kefaye, A., & Ashour, M. (2015). Effective diagnosis and monitoring of heart disease. *International Journal of Software Engineering and Its Applications*, 9(1), 143-156..
- [4] Nagavelli, U., Samanta, D., & Chakraborty, P. (2022). Machine Learning Technology-Based Heart Disease Detection Models. *Journal of Healthcare Engineering*, 2022.
- [5] Alarsan, F. I., & Younes, M. (2019). Analysis and classification of heart diseases using heartbeat features and machine learning algorithms. *Journal of big data*, 6(1), 1-15.
- [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] Arabasadi, Z., Alizadehsani, R., Roshanzamir, M., Moosaei, H., & Yarifard, A. A. (2017). Computer aided decision making for heart disease detection using hybrid neural network-Genetic algorithm. *Computer methods and programs in biomedicine*, 141, 19-26.
- [9] Centre for Disease Control and Prevention, Online available at: [http://www.cdc.gov/heartdisease/risk\\_factors.htm](http://www.cdc.gov/heartdisease/risk_factors.htm)
- [10] American Heart Association, Online available at: <http://www.heart.org/HEARTORG/Conditions>
- [11] Truong, V. T., Nguyen, B. P., Nguyen-Vo, T. H., Mazur, W., Chung, E. S., Palmer, C., ... & Le, T. K. (2022). Application of machine learning in screening for congenital heart diseases using fetal echocardiography. *The International Journal of Cardiovascular Imaging*, 1-9.
- [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.