IJDACR
ISSN: 2319-4863

International Journal of Digital Application & Contemporary Research
Website: www.ijdacr.com (Volume 7, Issue 03, October 2018)

Prediction of Heart Disease using Major Risk Factors based on ABC Optimized Neural Network

Priyanka Mandot
M. Tech. Scholar
Department of Computer Science
Shrinathji Institute of Technology & Engineering,
Nathdwara (India)
priyankamandot@gmail.com

Ajay Singh Dhabariya
Assistant Professor
Department of Computer Science
Shrinathji Institute of Technology & Engineering,
Nathdwara (India)
ajay.itengineer@gmail.com

Abstract – **This paper is particularly interested in the classification of data. The classification allows obtaining a prediction model from training data and test data. These data are screened by a classification algorithm that through the combination of mathematical tools and computer methods produces a new model capable of classified data possibly having the same classes of data.**

The analysis of data in the field of medicine is becoming more frequent in order to clarify the diagnoses, to refine the research methods and to envisage appropriate supplies of equipment according to the importance of the pathologies that appear. To analyze the present data in order to predict optimal results. This paper aims to implement a framework for prediction of heart disease using major risk factors based on the Artificial Bee Colony (ABC) optimized Neural Network. Performance of proposed research work is evaluated using the true positive rate, false positive rate, accuracy, and specificity.

Keywords – **Artificial Bee Colony, Cardiovascular disease, MLP, Neural Network, RBF.**

I. Introduction

Cardiovascular disease (CVD) is the leading cause of death and disability in developed countries. They represent in the world, a major stake of public health [1]. In the terminology of CVD, the medical literature refers to various chronic pathologies or events that share a pathophysiology related to atherosclerosis and responsible for premature death. It can act of [2]:

- Coronary heart disease (stroke angina, unstable angina, myocardial infarction, sudden death).
- Cerebrovascular accidents (hemorrhagic or ischemic, transient or formed).
- Peripheral vascular diseases (peripheral arterial disease, aortic aneurysm, renal failure with nephroangiosclerosis).
- Heart failure

The socio-economic changes observed during the last decades have improved the living conditions in several countries of the world [3]. This improvement, combined with an increase in sedentary lifestyle and greater accessibility to high-fat foods, has been accompanied by an increase in CVD risk factors. Thus, in some countries, half of diabetic patients die from CVD [4].

This is partly due to a significant increase in the prevalence of various risk factors associated with these diseases (obesity, insulin resistance, type 2 diabetes, dyslipidemia, high blood pressure) in adults and adolescents. Moreover, a tendency to group together several of these factors in the same individual has been observed, giving rise to the concept of a metabolic syndrome with a high cardiovascular risk [5].

Acute ischemic attacks (sudden death, myocardial or cerebral infarction, and unstable angina) are, in the majority of cases, the clinical translation of atheromatous disease, following rupture or erosion of the plaque of atherosclerosis with formation of an atherosclerotic plaque [6].

II. Cardiovascular Risk Factors

A cardiovascular risk factor is defined as a factor in which the patient's exposure to this factor increases the risk of developing a CVD, while the removal or improvement of this factor decreases this risk [7]. The importance of the risk factor considered is defined by the strength of the association with the disease (expressed by the relative risk observed in the exposed subjects compared to the unexposed), and / or the gradual association (parallel to the level of the risk factor). This definition implies a notion of causality between the factor and the disease [8]. This means that the risk factor precedes the disease (notion of anteriority) and that the correction of the factor must make it possible to reduce the incidence

IJDACR
International Journal Of Digital Application & Contemporary Research

International Journal of Digital Application & Contemporary Research
Website: www.ijdacr.com (Volume 7, Issue 03, October 2018)

of the disease (concept of reversibility). It must of course be recognized in several different populations and offer a plausible physiopathological explanation of the disease [9].

Strictly speaking, when there is no direct causal relationship, it is actually a "risk marker", a witness to a process (e.g., elevation of microalbuminuria, elevation of C-reactive protein "CRP"). While there is a direct causal link between the agent and the disease, it is a genuine risk factor [10].

Overall cardiovascular risk is defined as the probability of developing a cardiovascular event in a given time depending on several risk factors considered simultaneously [11]. Born more than 50 years ago with the advent of the analytical epidemiology of cardiovascular diseases, this concept now appears as a potential prevention tool integrating the multifactorial nature of these diseases [12].

A. *Physiological Factors*

1) *Age*
2) *Sex and Menopausal Status*

B. *Lifestyle Factors*

1) *Smoking*
2) *Physical Activity*
3) *Alcohol*

III. Proposed Method

A. *Data Analysis and Encoding*

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 [13]-[14]. 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, and obesity. The system indicated whether the patient had risk of heart disease or not.

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

B. *Neural Network*

Among the different architectures of artificial neural networks we adopted for this study the Multi-Layer Perceptron (MLPs) [15]. MLPs are most commonly used in supervised learning approaches, that is, when an association between two types of data, representing network input and output respectively, must be learned. In an MLP artificial neurons are organized in layers.

Let $y = f(x)$ be a function with $x \in R^D$ and $x \in R^D$ and let $H_i, i = 1, \dots N$ a set or basic functions.

The function f can be written in the form (1):

$$y = f(x) = \hat{f}(x) + r(x) \qquad (1)$$

Where $r(x)$ is the residue.

The function $f(x)$ can be approximated by $\hat{f}(x)$ given by the form (2)

$$y \simeq \hat{f}(x) = \sum_{i=1}^{N} w_i h_i(x) \qquad (2)$$

The goal is to minimize the error by adjusting the w_i settings appropriately. A possible choice for the approximation error is the L_2 standard of the residual function $r(x)$ is defined as:

$$(\|r(x)\|L_2)^2 = \int r(x)^2 dx \qquad (3)$$

Approximation of functions by RBF networks. The output of an RBF network is given by equation (4):

$$\hat{y} = \hat{f}(x) = \sum_{i=1}^{N} a_i \emptyset_i(x, \mu_i, \sigma_i) \qquad (4)$$

Using the above equation (4), the function $\hat{f}(x)$ can be written as:

$$y = \sum_{i=1}^{N} a_i \emptyset_i(x, \mu_i, \sigma_i) + r(x) = \hat{y} + r(x) \quad (5)$$

The error can be minimized by appropriately adjusting the weights a_i, centres μ_i and the widths σ_i.

When the function $f(x)$ is unknown, and having a set of couples outputs (x_i, y_i), $i = 1 \dots . n$,the RBF network can be built and it can be learn it to follow the function $f(x)$.

Learning this network is an optimization problem to choose the values of centres, weights and widths to minimize criterion:

$$J^2(\mu, \sigma, a) = \sum_{i=1}^{n} \|y_i - \hat{y}\|^2{}_i \qquad (6)$$

Learning approach of RBF networks by the algorithm ABC optimized learning neural networks with radial base functions with N neurons can be considered as an optimization problem whose objective is to reduce the error between the desired output of the function to be approximated and the output of the network.

C. *Artificial Bee Colony Optimization of Neural Network*

The ABC algorithm was developed by Karaboga and Basturk in 2007 [16], inspecting the behaviours of real bees to find the food source, called nectar, and sharing information from food sources to others bees in the nest.

In this algorithm, artificial bees are defined and classified into three groups, employing bees (bees seeking food), spectators (observation bees) and scouts (scouts) are responsible for finding new foods (nectar from new sources).

If the honey bee from a site fails to find the source of food, it must necessarily become a scout to randomly search for new sources of food. Employer

IJDACR
International Journal Of Digital Application & Contemporary Research

International Journal of Digital Application & Contemporary Research
Website: www.ijdacr.com (Volume 7, Issue 03, October 2018)

bees share the information with the bees in a beehive so that the honey bees can choose a food source to explore.

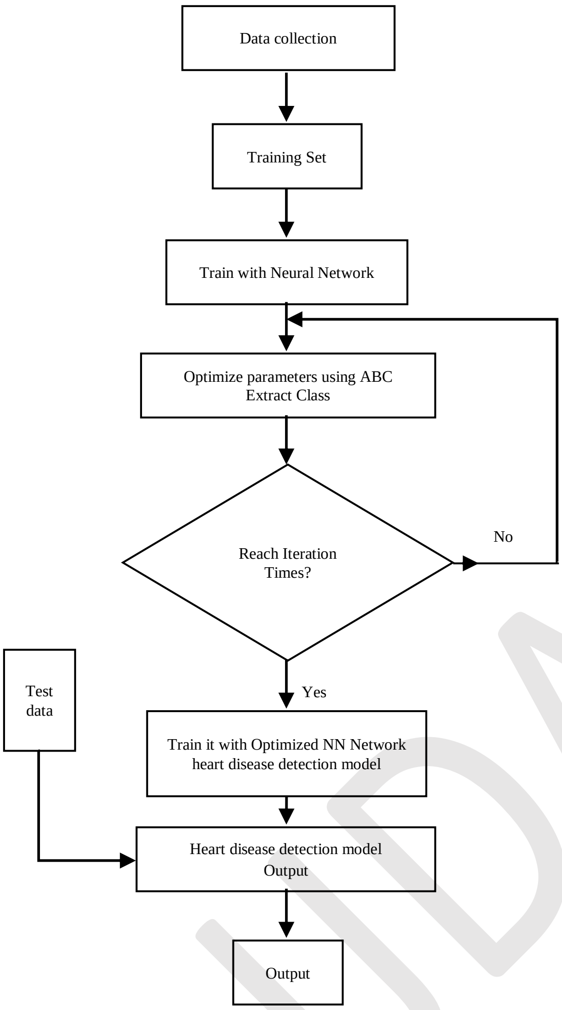

Figure 1: Flow diagram for ABC optimized Neural Network based heart disease detection model

The process of the ABC algorithm is presented as follows:

Step 1- Initialization:

Start by selecting Fe percentage of population randomly in the search space using the following equation:

$$U_j = U_j^{min} + n_j * \left(U_j^{max} - U_j^{min}\right) \quad n_j \in [0,1] \tag{7}$$

Knowing that each bee carries a vector "U" of "n" solution.

Assess them using the objective function, and calculate their Fitness values, called the amount of liquid by the following equation:

$$Fitness = \frac{1}{F_{objective}+1} \tag{8}$$

Fe represents the ratio of bees in the total population.

Once these populations are placed in the search space, they take the name of the employing bees.

Step 2- Movement of Employer Bees:

Calculate the probability of choosing a food source by the equation:

$$P_i = \frac{0.9*Fitness_i}{\max(Fitness_i)} + 0.1 \tag{9}$$

Select a food source and determine its amount of nectar. The equation of motion of the observing bees is given as follows:

$$m_{ij}(t+1) = x_{kj} + y\left(x_{ij}(t) - x_{kj}(t)\right) \quad y \in [0,1] \tag{10}$$

x_{ij} is the number of iterations, xij is the honeybee used randomly, j is the size of the solution vector that produces a series of random variables in the range [-1.1]; where k∈ {1,2,3, ... N} and j∈ {1,2, ... D} are chosen randomly with 'k' different from 'i', 'D' is the number of parameters to optimize.

As m_{ij} is the i^{th} position of the bee spectator, t is the number of iterations, x_{ij} is the bee used randomly, $'j'$ represents the size of the solution vector which produces a series of random variables in the range [-1,1]; where $k \in \{1,2,3, \dots N\}$ and $j \in \{1,2, \dots D\}$ are chosen randomly with $'k'$ different from $'i'$, $'D'$ is the number of parameters to optimize.

Step 3- Move Scouts:

If the fitness values of the employing bees are not improved by a predetermined number of iterations, these food sources are dropped, and the bee found in this location will randomly move to explore other new locations. (Employer bees become Scouts). This explanation is mathematically translated by the equation:

$$V_{ij} = V_{ij}^{min} + \varphi_{ij} * \left(V_{ij}^{max} - V_{ij}^{min}\right) \quad \varphi_{ij} \in [0,1] \tag{11}$$

Step 4- Update the best food source found so far:

Learn the best fitness value and position, which are found by the bees, and memorize them.

Step 5- Stop Criterion:

Check the calculation process until the number of iterations reaches the predefined maximum value or a solution of the acceptable objective function is found.

The fitness function used is mean squared error (MSE) which is calculated as below:

$$mse = \sum_k (O_k - T_k)^2 / n \tag{12}$$

IV. SIMULATION AND RESULTS

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

IJDACR
ISSN: 2319-4863

IJDACR
International Journal Of Digital Application & Contemporary Research

International Journal of Digital Application & Contemporary Research
Website: www.ijdacr.com (Volume 7, Issue 03, October 2018)

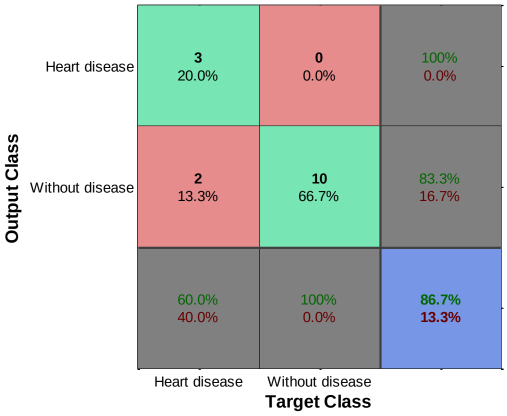

Figure 2: Confusion matrix for NN method

Figure 3: Confusion matrix for ABC optimized NN method

V. CONCLUSION

Classification in data mining ensures a reduction in the size of the problem, which reduces the duration of learning and simplifies the learned model. This simplification generally facilitates the interpretation of this model. It also makes it possible to avoid the phenomenon of over-learning improving the accuracy of the prediction and the understanding of the classifier. In this paper, Artificial Bee Colony algorithm optimizes the link weights of radial basis function and then given as input to the 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 ABC-Optimized NN approach outperforms the NN based approach.

REFERENCE

[1] Cífková, R. and Krajčoviechová, A., 2015. Dyslipidemia and cardiovascular disease in women. *Current cardiology reports*, *17*(7), p.52.

[2] Daljeet Kaur, Aman Paul, "Performance Analysis of Different Data mining Techniques over Heart Disease dataset", International Journal of Current Engineering and Technology, Vol. 4, No. 1, February 2014.

[3] B. Venkatalakshmi, M. V. Shivsankar, "Heart Disease Diagnosis Using Predictive Data mining", International Journal of Innovative Research in Science, Engineering and Technology (IJIRSET), Volume 3, Special Issue 3, March 2014.

[4] Nilakshi P. Waghulde, Nilima P. Patil, "Genetic Neural Approach for Heart Disease Prediction", International Journal of Advanced Computer Research, Vol. 4, No. 3, Issue 16, September 2014.

[5] Shen, B.J., Todaro, J.F., Niaura, R., McCaffery, J.M., Zhang, J., Spiro III, A. and Ward, K.D., 2003. Are metabolic risk factors one unified syndrome? Modeling the structure of the metabolic syndrome X. *American journal of epidemiology*, *157*(8), pp.701-711.

[6] Reed, G.W., Rossi, J.E. and Cannon, C.P., 2017. Acute myocardial infarction. *The Lancet*, *389*(10065), pp.197-210.

[7] Hlaudi Daniel Masethe, Mosima Anna Masethe, "Prediction of Heart Disease using Classification Algorithms", IEEE Proceedings of the World Congress on Engineering and Computer Science San Francisco, USA, Vol. 2, October 2014.

[8] K. Thenmozhi, P. Deepika, "Heart Disease Prediction Using Classification with Different Decision Tree Techniques", International Journal of Engineering Research and General Science Vol. 2, Issue 6, November, 2014.

[9] Monika Gandhi, Dr. Shailendra Narayan Singh, "Predictions in Heart Disease Using Techniques of Data Mining", IEEE, 1st International Conference on Futuristic trend in Computational Analysis and Knowledge Management (ABLAZE), pp. 520-525, February 2015.

[10] R. Tamilarasi, Dr. R. Porkodi, "A Study and Analysis of Disease Prediction Techniques in Data Mining for Healthcare", International Journal of Emerging Research in Management &Technology, Vol. 4, Issue 3, March 2015.

[11] Ankita Dewan, Meghna Sharma, "Prediction of Heart Disease Using a Hybrid Technique in Data Mining Classification", IEEE, 2nd International Conference on Computing for Sustainable Global Development, pp. 704-706, March 2015.

[12] G. Purusothaman, P. Krishnakumari, "A Survey of Data Mining Techniques on Risk Prediction: Heart Disease", Indian Journal of Science and Technology, Vol. 8, Issue 12, June 2015.

[13] Centre for Disease Control and Prevention, Online available at: http://www.cdc.gov/heartdisease/risk_factors.htm.

[14] American Heart Association, Online available at: http://www.heart.org/HEARTORG/Conditions

[15] Tang, J., Deng, C. and Huang, G.B., 2016. Extreme learning machine for multilayer perceptron. *IEEE transactions on neural networks and learning systems*, *27*(4), pp.809-821.

[16] Karaboga, D. and Basturk, B., 2007. A powerful and efficient algorithm for numerical function optimization: artificial bee colony (ABC) algorithm. *Journal of global optimization*, *39*(3), pp.459-471.