

A Optimized Binary Approach for Heart Disease Decision Model in Biomedical Data Mining

Sonika Rana
Department of computer
science And Engineering
Sachdeva Engineering
college for girls ,Gharuan (Mohali) PB

Sukhjot Kaur
Department of computer
science And Engineering
Sachdeva Engineering
college for girls ,
Gharuan (Mohali) PB

ABSTRACT

Heart diseases are the most common cause of death across worldwide. Due to the huge amount of data obtained from medical sector, hidden patterns remain hidden and knowledge cannot be extract from the database. This will result into loss of required information and treatment cannot be done accordingly. Several techniques have been proposed till now to find the effect of disease at earlier stage but still it is under consideration. Data mining is used to extract useful information from the database. Consequently, data analyses tool can be used to fetch required information from the database and effective decision can be made. Traditionally various heart disease detection techniques have been proposed like decision tree, genetic algorithm and so on.

Keywords

Data Mining, Heart Disease, BPSO, Neural Network.

1. INTRODUCTION

Data mining is a process which comes under the category of computer science in order to investigate large data set which belongs to a pattern. Here large data set stands for Big Data. The pattern in big data contains the method or techniques at the intersection of AI i.e. Artificial Intelligence and database systems or statistics etc. the aim of the data mining is to extract those part of data from data warehouse which is valid or meaningful and can be helpful in various decision making processes. Because the whole of the data stored in the data warehouse is not meaningful or important for each and every purpose. Data mining is a knowledge discovery process. It also performs data management and pre-processing, visualization, complexity consideration and online updating [21].

Health care sector is the most trending sector where there is a requirement for data mining to extract important, meaningful or required information. Data mining techniques are getting famous gradually in the several fields like prediction of diseases in healthcare systems where medical treatment can be compared according to their symptoms and able to identify successful treatment for a particular disease [12]. Data mining is also helpful in keeping track of states along with risk of a disease to the patient. Thus quality of service is required in order to get meaningful data from the dataset. Basically it has been used to minimize the risk of fraud and abuse in medical area. Diagnosis and laboratory procedures may be painful as well as costly for the patient but in case of data mining with health care system it reduces overall cost which leads to popularity and easy decision making. Therefore, before implementation of data mining in health care some factors need to be considered as quality, time, resources, efforts and money. After considering these factors collaboratively healthcare organization will be able to make benefit out of it.

Apart from several advantages it also includes some considering issues like collection of data, raw data for data mining available from different resources and stores in the data warehouse solution [20]. There can be another problem of capturing complete and consistent data at start due to which quality of data may be reduced. At last data mining requires planning and preparation to get best results.

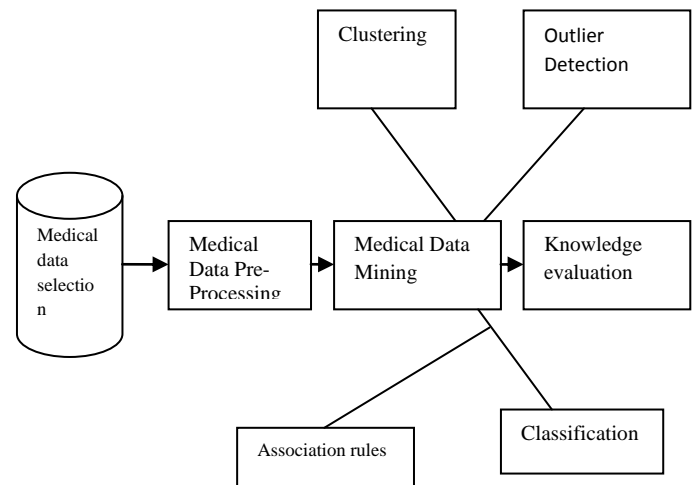


Figure 1. Data mining in medical sector

Advanced health disease prediction system assist answering to the questions arises with “what if” situation as compared to traditional decision support system. In the medical database, patient’s data is stored regarding its age, sex, blood pressure and other symptoms which help in prediction of a heart disease. Cause it provides knowledge regarding patterns and their relationship between medical factors [23].

With the increment in technology, rate of diseases is also increased. Heart attack diseases are the main cause of death in several countries but early detection can prevent this attack and patient can survive longer. In health disease prediction, data gathers from medical resources where hidden information is also present in them. Firstly, raw data is converted into dataset on which modeling has been performed using several data mining techniques and then decision is made on the acquired information.

From all the diseases, cardiovascular disease or CVD is the major cause of death worldwide. It has been researched that 50% of all deaths in Europe due to heart diseases and it is still increasing because of increasing elderly population. Along with cardiovascular diseases, another diseases like IHD i.e. ischemic heart diseases having myocardial infarction and heart failure. These diseases are the challenging issues in healthcare systems. From 1 to 5 per 1000 people suffering

from heart failure. Even though advances have been performed to get better surveillance treatment but yet Heart failure diagnosis has been occurred 1.7 years in men and 3.2 years in women. As compared with cardiovascular diseases, vascular heart diseases are less frequent.

Table1. Death rate of coronary heart disease among age and sex, 2008 [27]

	All ages	<35	35-44	45-54	55-64	65-75	>75
Men	52,5 85	114	834	2,809	6,80 9	11,8 85	30, 14 1
Women	41,7 96	34	185	588	1,88 3	5,49 4	33, 61 2
Total	94,3 81	148	1,019	3,397	8,68 5	17,3 79	63, 75 3

2. BPSO

Initially Particle Swarm optimization (PSO) technique has been introduced by Eberhart and Kennedy in 1995. The base of this algorithm is started by the behavior of the birds, bee and school of fishes. Thus PSO follows simulation behavior of the biological bird folk. Algorithm includes unpredictable and graceful compositions of the birds in providing information to the population. This is the main idea behind developing the PSO so that information can be shared with the individuals of the population. PSO follows search criterion which is based on the population of the particles. This algorithm is totally based on the exchange of information among individuals in the population whereas individuals referred as particles and population is considered as swarm [29]. Each particle in the swarm adjusts its position according to the previous experience and proceeds to the best previous position in the swarm. By memorizing the best own position in the swarm helps an individual particle to establish its experience which can be implied in local search as well as in global search. Experience can also be gained from the neighboring experience or the experience of the whole swarm [27].

PSO basically divided into two parts of search such as global neighborhood and local neighborhood. Global neighborhood is defined where each particle move towards the best previous position or the best particle in the swarm known as gbest model [29]. On the other hand, if particle move towards its best previous position and towards the best particle in the restricted neighborhood in the swarm is known as lbest model.

Each particle in the swarm is represented through the vector in the form of multidimensional search space. Thus given vector is assigned with another vector which helps in determining next movement of that particle. Such process is known as velocity vector. Another property of the PSO algorithm concludes that velocity of a particle can also be update on the basis of current velocity and the best position explored by the particle as well as on the basis of gbest model explored by the

swarm. PSO can be applied on various areas like power and voltage control, task assignment, mass-spring system etc [1].

Furthermore, Kennedy and Eberhart proposed a technique based on PSO named as Binary PSO in 1997 to resolve the discrete binary problems. In case of binary problems two terms are considered as “yes or no”, “true or false”, “one or zero” and “include or not to include” etc. This will help in determining real value in binary search space. In binary PSO, real valued version can be updated through particle’s personal best and global best. Consequently, range will be defined under restricted dataset as [0, 1]. Thus a map is required to map all the real valued numbers of the velocity from [0, 1].

Normalized function is named as sigmoid function in this algorithm as:

$$V_{ij}(t) = \text{sig}(V_{ij}(t)) = \frac{1}{1 + e^{-V_{ij}(t)}} \dots \dots \dots (1)$$

To update the velocity vector of the particle equation 1 can be used and new position of the particle will be obtained after applying equation below:

$$X_{ij}(t+1) = \begin{cases} 1 & \text{if } r^{ij} < \text{sig}(V_{ij}(t+1)) \\ 0 & \text{otherwise} \end{cases} \dots \dots \dots (2)$$

Where r_{ij} defines uniform random number in the range [0, 1].

3. PROBLEM FORMULATION

Evaluation of a status of a person is difficult for the physician from huge amount of data. Normalized or traditional approach is not accurate enough that can provide the status of a person. Thus there is a requirement of a model that can process the large database and provides resultant data to the physician. In the traditional optimization algorithms that have been proposed in which a certain path selects by the insect defines the most appropriate path. Several traditional algorithms as Cuckoo optimization, Ant colony optimization, Bee colony optimization has been used for selection of an appropriate path. These techniques are helpful in making effective decision. But in certain cases they do not provide efficient results. Proposed technique provides accurate results in terms of fitness value. But in the traditional approach, accuracy is not up to the mark which leads the proposing of new model.

4. PROPOSED WORK

In our propose work we are taking the dataset of biomedical field that is of heart rate which is having entries the system which we are going to develop will surely provide us the accuracy more than the previous algorithms.

In our proposed methodology a swarm intelligence algorithm that is BPSO is used for optimizing the neural network on basis of Weight values it will update the results of weight value until and unless our system got more accurate results than the Initial results given by normal neural network.

4.1 Objectives

1. To develop an system which will optimized the system to get accuracy after testing.
2. To compare results of Normal Neural Network And Optimized Neural network
3. Use Swarm Intelligence algorithm for Optimization.
4. Developing fast and More accurate Proposed algorithm.

5. METHODOLOGY

Proposed work has been done to obtain better and efficient results and in terms of prediction. So that death rate due to the heart disease can be minimized. Methodology of the proposed work is given below:

1. Preliminary take predefined dataset of heart disease entries for manipulation purposes.
2. Secondly capture some of the data for training which will be performed through artificial neural network.
3. Training has been performed on rest of the data using neural network.
4. Now evaluate accuracy of the trained dataset and save them for the future comparison.
5. Extract trained data and evaluate weight values in terms of performance to individual data value.
6. Now using BPSO i.e. binary particle swarm optimization technique to update the Weight value and trained them again.
7. Test the whole system updated by the BPSO and obtained new neural network.
8. Obtain accuracy of the whole system to check the performance parameter.
9. Now checks trained dataset which was save at earlier stage and the results obtained after applying BPSO has been checked that whether the results obtained are optimized with respect to initial network without BPSO.
10. If NO then keep updating the weight value and repeat step from 6 to 8. On the other hand if YES then save the results and perform comparison between the optimized and normalized system.

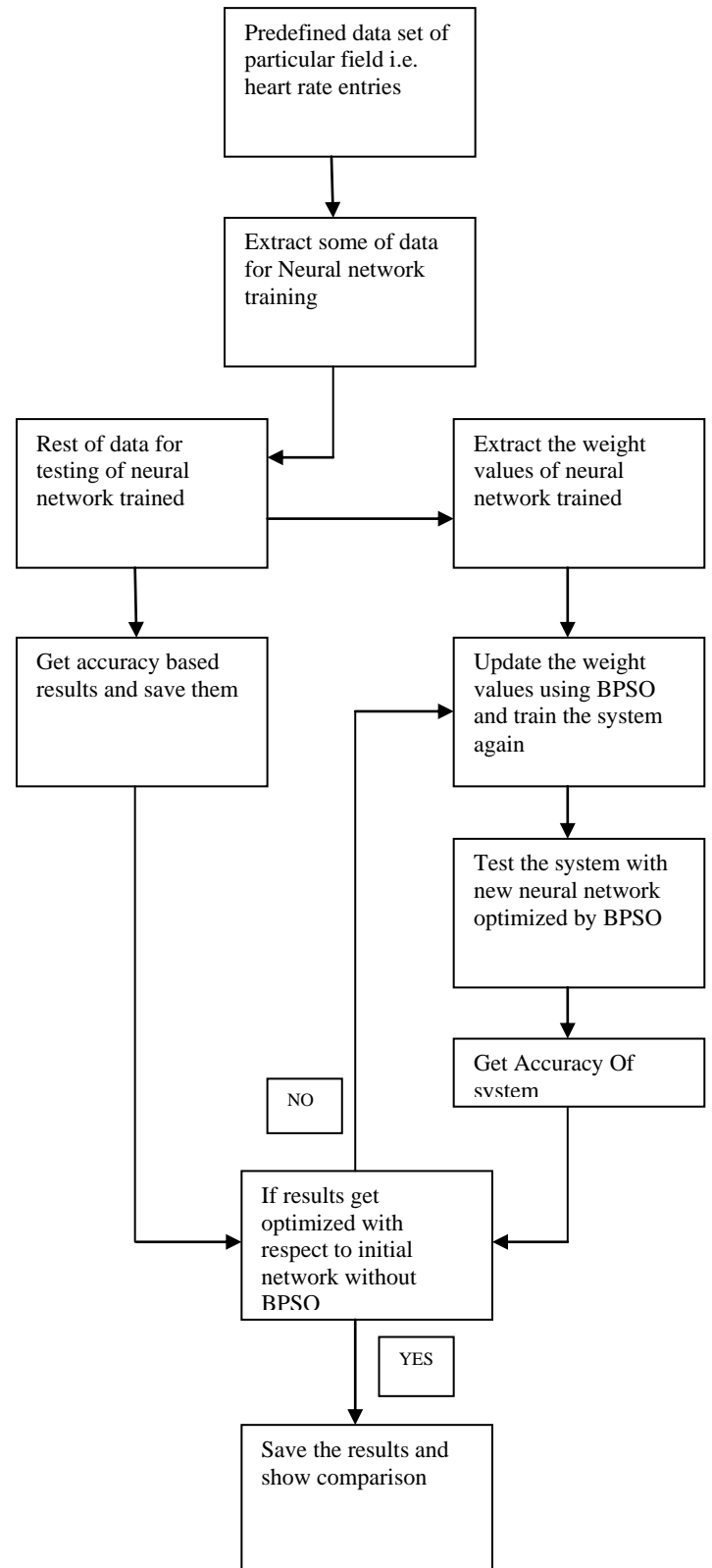


Figure 2. flow diagram of proposed work using Neuralnetwork and BPSO

6. RESULTS AND DISCUSSION

In this section proposed technique has been evaluated, comparison has been performed to proven given fact. As accuracy is the important and crucial factor of any technique. Thus proposed technique accuracy must be better. Results

have shown that enhanced techniques accuracy is better and efficient.

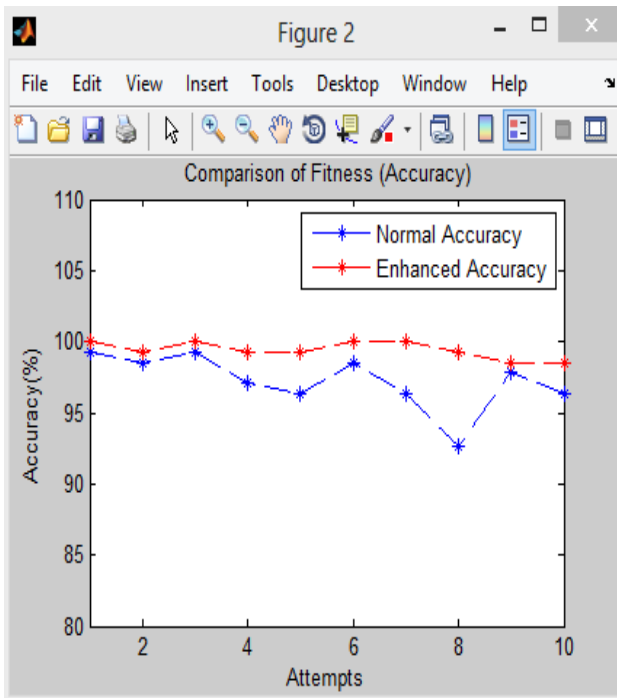


Figure3. Represents accuracy of normal and enhanced system

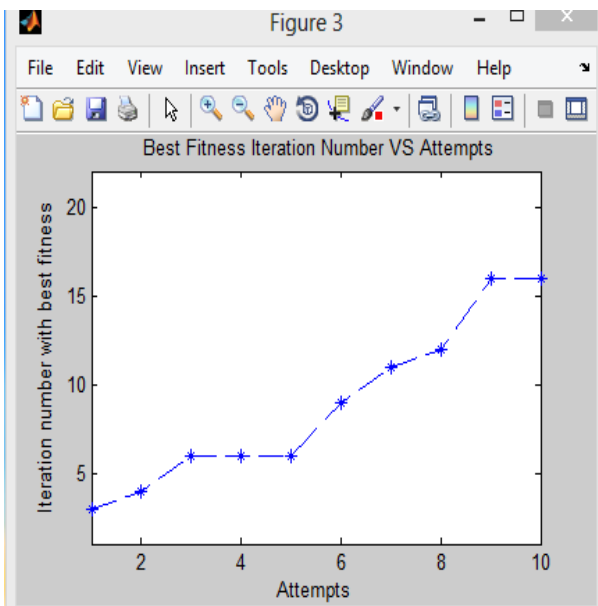


Figure4. Best fitness value acquired in number of attempts

Above graph represents number of iterations performed and acquired best fitness value in each iterations or attempts.

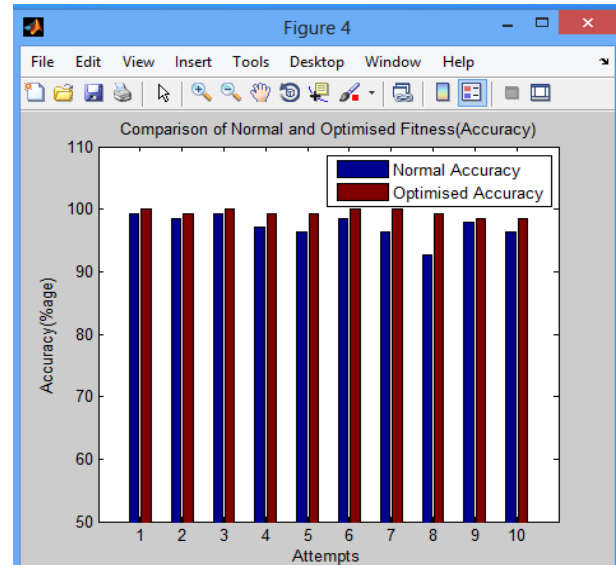


Figure5. Accuracy of normal and optimized system in terms of attempts made.

In the figure 3, it shows accuracy performed in individual attempts and demonstrated that optimized accuracy surpass normal accuracy.

7. CONCLUSION AND FUTURESCOPE

Medical sector contains huge amount of data. Physician may not analyze hidden patterns and process knowledge. Due to which data analysis tool can be used to process and can make effective decision. Heart disease is the main cause of death across worldwide. Thus, there is a requirement of reducing the risk of heart disease using data analysis tool. Experiments have done through which evaluates accuracy. Results have proven that normalized classification do not provide accuracy as compared to optimized classification. Moreover fitness value acquired from the optimized technique is far better than normalized. Proposed technique uses behavior of the insects as artificial PSO and neural network is combined which produces accurate results.

From the results it can be concluded that proposed technique is better in terms of accuracy and fitness value. Prediction of heart disease can be done at earlier stage using proposed method. In future more techniques can be enhanced in heart disease prediction which follows biological behavior and provides better results along with their fitness value.

8. REFERENCES

- [1] Kennedy, James. "Particle swarm optimization.", Springer, Pp. 760-766, 2011.
- [2] Umair Shafique et al, "Data Mining in Healthcare for Heart Diseases", International Journal of Innovation and Applied Studies, Vol. 10, No. 4, Pp. 1312-1322, Mar 2015.
- [3] Sellappan Palaniappan et al, "Intelligent Heart Disease Prediction System Using Data Mining Techniques", IEEE, Pp. 108-115.
- [4] Hlaudi Daniel Masethe et al, "Prediction of Heart Disease using Classification Algorithms", WCECS, Vol. 2, Pp. 22-24, October 2014
- [5] Dušan teodorović1 et al, "Bee colony optimization – a cooperative learning approach to complex transportation problems", Pp 51-60.

- [6] Mandeep Kaur Bedi et al, "Comparative Study of Two Natural Phenomena", *International Journal of Scientific & Engineering Research*, Vol.4, No. 3, March 2013
- [7] Elham Shadkam et al, "Evaluation The Efficiency Of Cuckoo Optimization Algorithm", *International Journal on Computational Sciences & Applications*, Vol.4, No.2, Pp. 39-47, April 2014
- [8] Swagatam Das et al, "Bacterial Foraging Optimization Algorithm: Theoretical Foundations, Analysis, and Applications
- [9] Latha Parthiban et al, "Intelligent Heart Disease Prediction System using CANFIS and Genetic Algorithm", *International Journal of Biological and Medical Sciences*, Pp. 157-160
- [10] Christian Blum, "Ant colony optimization: Introduction and recent trends", *ELSEVIER*, Pp. 353-373
- [11] Rafael S. Parpinelli et al, "Data mining with an Ant Colony Optimization Algorithm
- [12] T.Y. Chen et al, "Application of data mining in a global optimization algorithm", *ELSEVIER*, Vol. 66, Pp. 24-63, December 2013
- [13] K.Sudhakar et al, "Study of Heart Disease Prediction using Data Mining", *International Journal of Advanced Research in Computer Science and Software Engineering*, Vol. 4, No. 1, Pp. 1157-1160
- [14] H Bosma et al, "Two alternative job stress models and the risk of coronary heart disease", *American Journal of Public Health* January, Vol. 88, No. 1, pp. 68-74, 1988
- [15] Punam Bajaj et al, "Review on Heart Disease Diagnosis Based on Data Mining Techniques", *IJSR*, Vol. 3, No. 5, Pp. 1593-1596, May 2014
- [16] Deepali Chandna, "Diagnosis of Heart Disease Using Data Mining Algorithm", *IJCSIT*, Vol. 5, No. 2, Pp. 1678-1680, 2014
- [17] Jyoti Soni et al, "Predictive Data Mining for Medical Diagnosis: An Overview of Heart Disease Prediction", *IJCA*, Vol. 17, No. 8, Pp. 43-48, March 2011
- [18] Asha Rajkumar et al, "Diagnosis of Heart Disease Using Data mining Algorithm", Vol. 10, No. 10, Pp. 38-43, September 2010
- [19] M. ANBARASI et al, "Enhanced Prediction of Heart Disease with Feature Subset Selection using Genetic Algorithm", *IJEST*, Vol. 2, No. 10, Pp. 5370-5376, 2010
- [20] K.Srinivas et al, "Applications of Data Mining Techniques in Healthcare and Prediction of Heart Attacks", *IJCSE*, Vol. 2, No. 2, Pp. 250-255, 2010
- [21] Nidhi Bhatla et al, "An Analysis of Heart Disease Prediction using Different Data Mining Techniques", *IJERT*, Vol. 1, No. 8, Pp. 1-4, October 2012Heart
- [22] Chaitrali S. Dangare et al, "A Data Mining Approach for Prediction of Heart Disease Using Neural Networks", *IJCET*, Vol. 3, No. 3, December 2012
- [23] K. Srinivas et al, "Analysis of coronary heart disease and prediction of heart attack in coal mining regions using data mining techniques", *IEEE*, Pp. 1344-1349, Aug. 2010
- [24] Yanwei Xing et al, "Combination Data Mining Methods with New Medical Data to Predicting Outcome of Coronary Heart Disease", *IEEE*, Pp. 868-872, November 2007
- [25] Monika Gandhi et al, "Predictions in heart disease using techniques of data mining", *IEEE*, Pp. 520-525, February 2015
- [26] Asha Rajkumar et al, "Diagnosis of Heart Disease Using Data mining Algorithm", Vol. 10, No. 10, Pp. 38-43, September 2010.
- [27] Chen, Guo-chu et al, "Particle swarm optimization algorithm.", *information and control-shenyang*, Vol 34, No.3, 2005, Pp. 318.
- [28] Allender, Steven, et al. "Patterns of coronary heart disease mortality over the 20 th century in England and Wales: Possible plateaus in the rate of decline." *BMC public health*, Vol. 8, No.1, Pp. 1, 2008.
- [29] Shi, Y., "Particle swarm optimization: developments", *IEEE*, Vol. 1, pp. 81-86, 2001.
- [30] Khanesar et al, "A novel binary particle swarm optimization", pp. 1-6, *IEEE*, June 2007.