



## INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY

### Heart Disease Diagnosis System Based On Data Mining And Neural Network

Ms. Preeti Gupta<sup>\*1</sup>, Ms. Punam Bajaj<sup>2</sup>

<sup>\*1</sup> Research Scholar, <sup>2</sup> Assistant Professor Department of CSE, Chandigarh Engineering College, Mohali, Punjab, India

[preeti.gupta30@gmail.com](mailto:preeti.gupta30@gmail.com)

---

#### Abstract

Data mining techniques have been commonly in the clinical decision support system for prediction and diagnosis of various diseases with good accuracy. One of the most important applications of such system is in diagnosis of heart diseases because it is one of the leading causes of the deaths all over the world. Heart disease diagnosis is a compound task which requires both experience and knowledge. The traditional way of predicting heart diseases are ECG, Heart MRI etc. Nowadays the healthcare industry collects huge amount of healthcare data which, unfortunately, are not “mined” to discover hidden information for effective decision making. Advanced data mining techniques can help to solve this situation. Hence this paper presents a technique for diagnosis of heart diseases using heart rate data set. These techniques are neural network and swarm intelligence optimization. In this paper Swarm Intelligence Optimization i.e. Genetic Algorithm has been used and applied for optimizing the Neural Network.

**Keywords:** Data Mining, Diagnosis, Genetic Algorithm, Heart Diseases, Neural Network, Swarm Intelligence Optimization.

---

#### Introduction

Heart Diseases remain the biggest cause of deaths for the last two decades. Recently computer technology and machine learning techniques are used to develop software to assist doctors in making decision of heart disease in the early stage. The diagnosis of heart disease depends on clinical and pathological data. Heart disease prediction system can assist medical professionals in predicting heart disease status based on the clinical data of the patients. In biomedical field data mining plays an essential role for prediction of diseases. In biomedical diagnosis, the information provided by the patients may include redundant and interrelated symptoms and signs especially when the patients suffer from more than one type of disease of the same category. The physicians may not able to diagnose it correctly. Data mining with intelligent algorithms can be used to tackle the said problem of prediction in medical dataset involving multiple inputs. Now a day's Artificial Neural Network has been used for complex and difficult tasks. The neural network is generally trained from the historical data with the expectation that it will discover hidden dependencies and that it will be able to use them for predicting. Feed-forward neural networks trained by back-propagation have become a standard technique for classification and prediction tasks.

Medical data mining has great potential for exploring the hidden patterns in the data sets of medical domain. These patterns can be utilized for clinical diagnosis. However the available raw medical data are widely distributed and heterogeneous in nature. These data need to be collected in an organized form. This collected data can then be integrated to form a hospital information system. Data mining technology provides a user-oriented approach to novel and hidden patterns in the data.

The WHO has estimated 17.3 million people died from CVD in 2008, representing 30% of all global deaths. Of these deaths, an estimated 7.3 million were due to coronary heart disease and 6.2 million due to stroke [11]. Hence more careful and efficient methods of cardiac diseases and periodic examination are of high importance. Data mining is the solution to this problem. Hence different data mining techniques can be applied to diagnose the disease more accurately.

#### Data mining techniques

“Data Mining is a non-trivial extraction of implicit, previously unknown and potential useful information about data”[3]. In recent years, data mining has found its remarkable hold in every field including healthcare. Data mining extract the patterns in the process of knowledge discovery in the database.

As the dataset has grown in size and complexity, new emerging field of data mining provides new techniques and methods which help to analyze and understand huge amount of data. Data mining involves some common methods that are association rule learning, clustering, classification, regression, summarization and sequential pattern matching. Classification is one of the most important techniques of data mining as it is the process of finding a set of models (or functions) which describe and distinguish data classes or concepts.

In classification, inputs are given in the form of a set of data, called a training set, where each record composed of several fields or attributes. These attributes are continuous, arriving from an ordered domain, or categorical, arriving from an unordered domain. One of the attributes, called the classifying attribute, indicates the class to which each dataset belongs. The objective of classification is the method to build a model of the classifying attribute based upon the other attributes which are not from the training data set.

Data mining is an interdisciplinary field. For the implementation of classification, techniques from other disciplines may also be applied, such as neural networks, fuzzy or high performance computing which will lead to intelligent, low cost solution. Data mining system may also combine techniques from image processing, information retrieval, spatial data analysis, pattern recognition and signal processing depending on the sort of data to be mined or on the given data mining application. Artificial neural network is one of the widely used data mining. Artificial neural networks has some advantages such as it automatically allow arbitrary nonlinear relations between the independent and dependent variables, and allows all possible interactions between the dependent variables. Due to above said advantages of ANN the use of neural network technique is adopted for the classification of dataset. Neural network approach is proposed for the efficient classification of medical data. Neural network is trained using feed-forward algorithm. To increase the efficiency of classification process, parallel processing is executed at each node in the network. After training the neural network performance of the network is analyzed with various set of data. In this paper, we will use artificial neural network to determine the attributes which contribute more towards the diagnosis of heart diseases.

### Artificial Neural Network

The Artificial Neural Network (ANN) is a technique which is applied to solve data mining applications. Neural Network is a closely interconnected network, which offers rich structure providing some features of the biological neural network. Neural network provides an advantage to the user to implement parallel concept at each layer level. Neural Network structure consists of three layers: an input layers, one or more hidden layers and the output layer. Number of hidden layers and number of neurons in each layer strongly depends on the complexity of system studied. The general scheme of typical three-layered ANN architecture is given in fig.1. ANN learns from existing examples. Artificial neural network are trained based on two training methods. 1. Supervised Training- for supervised learning external teacher is available which provide the neural network input data and actual desired output. The network parameters are adjusted under the combined influence of training vector and error signal to get that desired output for a given specific input. 2. Unsupervised Training-Input data and computation function is given to the network and output is calculated. It does not require the desired output. Neural network is an iterative learning process in which input data cases are given to the network one at a time and weights associated with input interconnection are adjusted each time [8].

The advantages of Neural Network for classification are:

- Robustness is more in Neural Network because of the weights.
- Performance of the Neural Network is improved by learning. Even after applying the training sets this may continues.
- For better performance the use of Neural Network can be parallelized as specified above.
- Once the appropriate training has been performed, low error rate and high degree of accuracy has been achieved.
- In noisy environment Neural Networks are more robust.

### Overview of techniques used

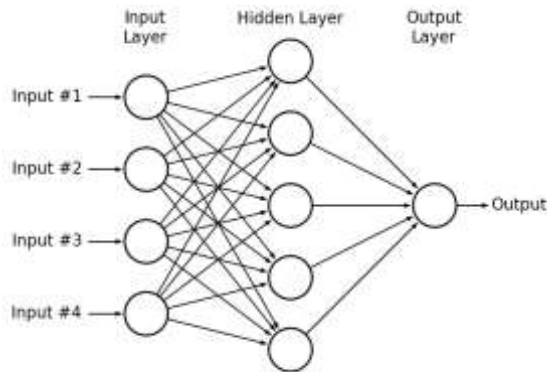


Figure: 1 General Structure of Neural Network

### Swarm Intelligence

Swarm intelligence is a computational intelligence technique which is used to solve compound real-world problems. It involves the collective study of behavior of individuals in a population who get across locally with one another and with their environment in a decentralized control system. The creativity often comes from nature, especially biological systems. The agent follows very simple rules and there is no centralized control structure to dictate how individual agents should act local and to a certain degree random, communication between such agents lead to the emergence of “intelligent” global behavior, unknown to the individual agents. Examples in natural systems of SI include that bird flocking, animal herding, bacterial growth, colonies and fish schooling. The definition of swarm intelligence is still not fully clear. In principle, it should be a multi-agent system that has self-organized behavior that shows some intelligent behavior [12].

### Genetic Algorithm

In the field of artificial intelligence, Genetic Algorithm is a search heuristic that impersonate the process of natural selection. This heuristic is commonly used to generate useful solution to optimization and search problem. Genetic Algorithm belongs to the larger class of Evolutionary Algorithms (EA), which generate solutions to optimization problems using techniques inspired by natural evolution, such as inheritance, mutation, selection and crossover.

A typical Genetic Algorithm requires:-

- A genetic representation of the solution domain
- A fitness function to evaluate the solution domain.

Once the genetic representation and fitness function are defined a Genetic Algorithm takes to initialize a population of solutions and then to improve it through recurrent application of the mutation, crossover, inversion and selection operators [13]. Basic steps in Genetic Algorithm are shown in Fig.2.

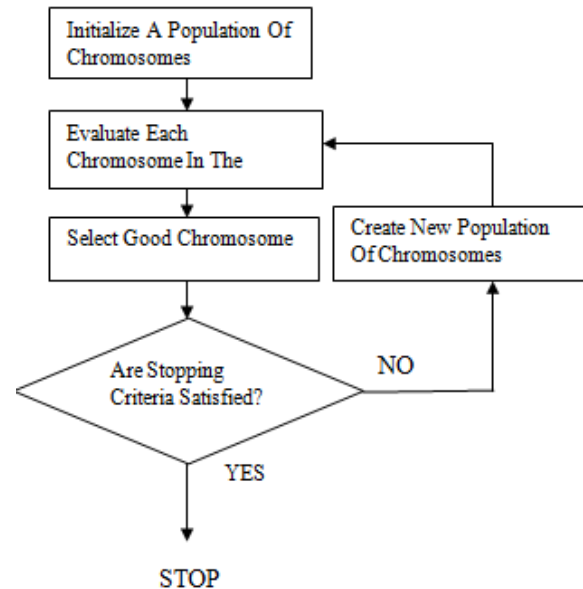


Figure.2 Flowchart Of Genetic Algorithm

### Literature survey

Work done in diagnosis of heart disease using data mining and artificial neural networks are discussed below:-

To predict heart disease, blood pressure and sugar, a system is proposed that uses neural network [7]. In this set of 78 records with 13 attributes are used for training and testing. To supervised network i.e. neural network with back propagation algorithm is used for training and testing of data.

**Heon Gyu Lee** [4] proposed HRV (Heart Rate Variability) technique to develop Multi-Parametric feature. They have used several classifiers to achieve this for e.g. Bayesian classifiers, C4.5 (Decision Tree) and SVM (Support Vector Machine), CMAR (Classification Based On Multiple Association Rules). **Sellappan Palaniappan** developed an Intelligent Heart Disease Prediction System (IHDPS) using data mining techniques, viz, Decision Trees, Naïve Bayes and Neural Network [9]. Each method has its own potency to get appropriate results. To build this system hidden patterns and relationship between them is used.

**Latha Parthiban** [5] formulated an approach for the prediction of heart disease on the basis of coactive neuro-fuzzy inference system (CANFIS). The

CANFIS model uses neural network capabilities with the fuzzy logic and genetic algorithm.

Decision support system for heart disease based on support vector machine and artificial neural network is proposed by **Gudadhe, Wankhede and Dongre**. For heart disease classification this paper proposed a decision support system based on Support Vector Machine and MLP neural network architecture. Support Vector Machine classifies the heart disease data into two classes which shows presence or absence of heart disease with 80.41% accuracy. Artificial Neural Network classifies the data into 5 categories of heart disease with 97.5% accuracy. This shows that both the methods gives high accuracy to classify the data but Artificial Neural Network classifies the data more accurately as compared to Support Vector Machine [6].

**Chaitali S.Dangare** [2] used neural network to predict heart disease with nearly 100% accuracy, by adding 2 more parameters i.e. Smoking and Obesity for better accuracy.

Cardio Vascular Disease prediction System using Neural Network and Genetic Algorithm is proposed by **Amma, N.G.B.** In this Genetic based Neural Network is used for training the system. The neural network final weights are saved in the weight base and are used for predicting the risk of cardio vascular disease. The accuracy of classification obtained using this approach is 94.17% [1].

Heart Disease Prediction system using weight optimized neural network is proposed by **T.Manju, K.Priya, R.Chitra**. This paper presents the application of Multi Layer Feed Forward Neural Network that integrates Genetic Algorithm and Back Propagation network for heart prediction. GA is used to initialize and optimize the connection weights of MLFFNN. The optimized NN is trained and tested using 270 patient data [10].

### Proposed methodology

The diagnosis process consists of two succeeding steps i.e. training and testing. The word Training means finding the neural network weights that minimize some error value whereas Testing means to evaluate the neural network with the best weights establish during training, using some measure of accuracy. When performing analysis with a set of existing data, split the data set into 70 % – 30 % for training and testing purpose respectively. The objective of this study is to measure the accuracy of this neural network and compare it with accuracy of optimized neural network which is optimized using swarm intelligence algorithm. Fig.3 depicts the flowchart of the work which is going to be carried out in our research.

Steps which will be followed in our research are as follows:-

- I. To develop a system and to optimize the system to get more accuracy after testing.
- II. To compare results of normal neural network and optimized neural network.
- III. Swarm intelligence algorithm i.e. Genetic Algorithm will be used for optimization.
- IV. To develop better and more accurate proposed system.

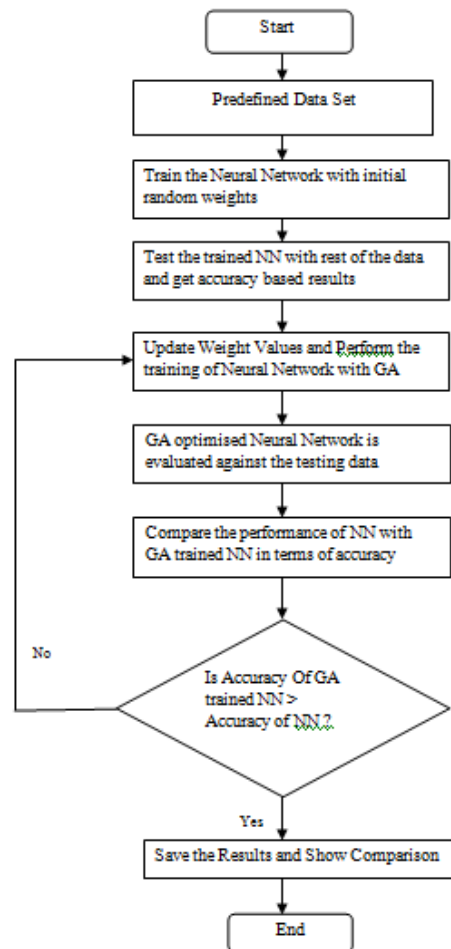


Figure: 3 Basic flowchart depicting training of Neural Network with GA

### Conclusion

In this paper, we have discussed that how different types of data mining techniques are used for diagnosis of heart diseases and also studied that how these techniques have performed better results when applied on different data sets. Each technique is unique in its own way, which might be suitable for different applications. Hence neural network can be effectively used in Heart Diseases Diagnosis. In our

future work, we will be training the neural network by using Swarm Intelligence Optimization i.e. Genetic Algorithm to optimize the system to get better results and more accuracy than the normal neural network.

### Reference

1. Amma, N.G.B “Cardio Vascular Disease Prediction System using Genetic Algorithm and Neural Network”, *IEEE International Conference on Computing, Communication and Applications*, 2012.
2. Chaitrali S Dangare “Improved Study Of Heart Disease Prediction System Using Data Mining Classification Techniques”, *International Journal Of Computer Applications*, Vol.47, No.10 (June 2012).
3. Frawley, G. Piatetsky-Shapiro, Christopher J. Matheus “Knowledge Discovery in Databases: An Overview”. Published by the *AI Magazine Volume 13 Number 3* (1992).
4. Heon Gyu Lee, Ki Yong Noh, Keun Ho Ryu, “Mining Biosignal Data: Coronary Artery Disease Diagnosis using Linear and Non linear Features of HRV”, *Emerging Technologies in Knowledge Discovery and Data Mining*, Vol 4819, 2007, pp 218-228.
5. Latha Parthiban and R. Subramanian, “Intelligent Heart Disease Prediction System using CANFIS and Genetic Algorithm”, *International Journal of Biological Biomedical and Medical Sciences*, Vol.3. No.3, 2008.
6. M. Gudadhe, K. Wankhade and S. Dongre, “Decision support system for heart disease based on support vector machine and artificial neural network”, *IEEE International Conference on Computer and Communication Technology*, pp. 741-745, September 2010.
7. Niti Guru, Anil Dhaiya, Navin Rajpal, “Decision Support System for Heart Disease Diagnosis using Neural Network”, *Delhi Business Review*, Vol.8, No.1 (January-June 2007).
8. R. Rojas, (1996) “Neural Networks: a systematic introduction”, Springer-Verlag.
9. Sellappan Palaniappan, Rafiah Awang, “Intelligent Heart Disease Prediction System using Data Mining Techniques”, *IJCSNS International Journal of Computer Science and Network Security*, Vol.8 No.8, August 2008.
10. T. Manju, K. Priya, R. Chitra “Heart Disease Prediction System Using Weight Optimized neural Network”, *International Journal Of Computer Science and Management Research*, Vol 2, Issue 5 May 2013
11. [cardiovascular\\_diseases/about\\_cvd/en/](http://cardiovascular_diseases/about_cvd/en/).
12. [en.wikipedia.org/wiki/swarm\\_intelligence](http://en.wikipedia.org/wiki/swarm_intelligence).
13. [en.wikipedia.org/wiki/Genetic\\_algorithm](http://en.wikipedia.org/wiki/Genetic_algorithm).