

## Implementation of Datamining to Predict the Heart Disease

**Tejas Anant Sapkal<sup>1</sup>, Prasad Deoram Pingale<sup>2</sup>, Premraj Manoj Ingawale<sup>3</sup>, Vikas Baban Pagre<sup>4</sup>, Komal Jagdale<sup>5</sup>**

*<sup>1,2,3,4,5</sup> Department of Computer Engineering , ISB&M School of Technology, Pune, India*

---

### Abstract

---

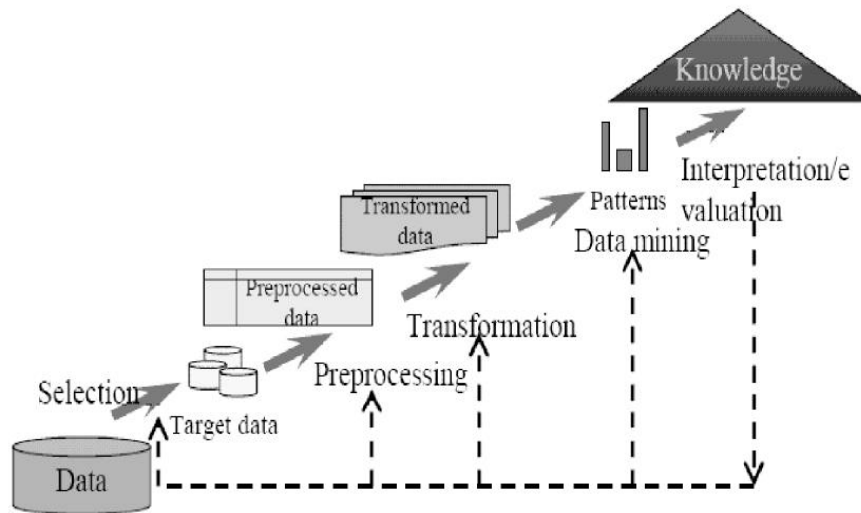
Today medical field have come a much extended way to treat patients with various kind of diseases. Among the most disturbing one is the Heart disease which cannot be seen with a naked eye and comes immediately when its limits are reached. Bad clinical decisions would cause death of a patient which cannot be able to spare or give up by any hospital. To attain an accurate and cost effective dealing computer-based and support systems can be developed to make correct decision. A lot of hospitals use clinical information systems to manage their healthcare or patient information. These systems produce huge amounts of data in the form of charts , images, text, and numbers. This data is rarely used to support the medical decision making. There is an amount of hidden information in this data which is not yet explored which give rise to the important problem of how to make useful information out of the data.

**Keywords:** Back-Propagation; Neural Network; Genetic Algorithm

---

### 1. Introduction

Every people desire to live a very luxurious life so they work similar to a machine in order to earn lot of money and live a relaxed life so in this race they forget to take care of themselves, because of this there food habits change their entire lifestyle change, in this type of lifestyle they are more tensed they have blood pressure, sugar at a very young age and they don't give enough rest for themselves and eat what they get and they even don't bother about the quality of the food if sick the go for their own medication as a result of all these small negligence it leads to a major threat that is the heart disease Data Mining is the exploration of large datasets to extract hidden and previously unknown patterns, relationships and knowledge that are difficult to detect with traditional statistics. Data mining techniques are the result of a long process of research and product development. The key functionality used in data mining are classification, association and clustering.



**Fig 1: The transition from raw data to valuable knowledge**

The data mining involve a small number of steps from raw data collection to some form of new information. The iterative procedure consists of steps like *Data cleaning*, *Data Integration*, *Data Selection*, *Data transformation*, *Data Mining*, *Pattern Evaluation*, and *Knowledge Representation*. The Figure 1 demonstrates that the Data Mining is the core for Knowledge Discovery Process The heart is important organ of human body part. It is nothing more than a pump, which pumps blood through the body. If circulation of blood in body is inefficient the organs like brain suffer and if heart stops working altogether, death occurs within minutes. Life is completely dependent on efficient working of the heart. The term Heart disease refers to disease of heart & blood vessel system within it. According to survey conducted by register general of India, heart disease is a major cause of death in India.

***Various kind of heart diseases are***

Coronary heart disease, Cardiomyopathy, Cardiovascular disease, Ischemic heart disease, Heart failure, Hypertensive heart disease, Inflammatory heart disease, Valvular heart disease

***Common risk factors of heart disease include***

High blood pressure, Abnormal blood lipids, Use of tobacco, Obesity, Physical inactivity, Diabetes, Age, Gender, Family history

As per the review conducted by WHO, out of 10 deaths in India, eight are caused by *cardiovascular diseases and diabetes*. Preventive strategies to reduce risk factors are essential and to decrease the alarmingly increasing burden of heart disease in our population.

**2. Literature review**

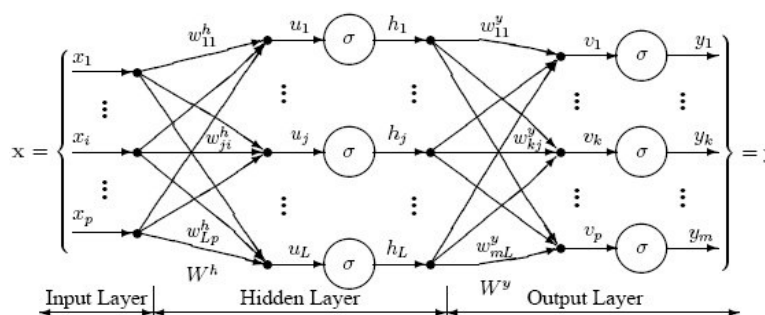
*Khaing (2011)* presented an efficient method for the forecast of heart attack risk levels from the heart disease database. Firstly, the heart disease database is clustered by the *K-means* clustering algorithm, which will extract the data relevant to heart attack from the database. They have employed the *ID3 algorithm* as the training algorithm to show level of heart attack with the decision tree. The results showed that the designed prediction system is able of predicting the heart attack effectively. In year 2012 Mai Shouman, et. al. carry out a work Applying k-Nearest

Neighbour in Diagnosing Heart Disease Patients. In this manuscript the author detailed outwork that applied *KNN* on a Cleveland Heart Disease dataset to investigate its efficiency in the prediction of heart disease. The author also investigated if the accuracy could be enhanced by integrating voting with *KNN*. The results show that applying voting could not enhance the *KNN* accuracy in the diagnosis of heart disease. S. Vijayarani et. al. (2013) performed a work on Efficient Classification Tree Technique for Heart Disease Prediction. This paper analyzes the classification tree methods in data mining. The classification tree algorithms used and tested in this work are *Decision Stump*, *Random Forest*, and *LMT Tree algorithm*. The objective of this research was to compare the outcomes of the performance of different classification techniques for a heart disease dataset. This work is done by using Waikato Environment for Knowledge Analysis (WEKA).

### 3. Methodology

A genetic algorithm is one of the search technique used in computing to find true or approximate results to optimization and search problems. Genetic algorithms are grouped as global search heuristics. Genetic algorithms are appropriate class of mutative algorithms that use methods motivated by developmental biology such as inheritance, variation, collection, and traverse (also called recombination). Genetic algorithms are implemented as a computer simulation in which a population of hypothetical representations (called chromosomes or the genotype or the genome) of candidate solutions (called individuals) to an optimization problem evolves to form a better results. The evolution usually starts from a population of arbitrarily generated by every person and takes place in generations. In each generation, the fitness of every person in the population is calculated, multiple individuals are selected from the current population (fitness value), and modified population (recombined) to form a new population

*Back Propagation Neural Network* is the one of the more focused area of research. The basic architecture consists of input, hidden and output layers. In this the weights are updated according to the error at the output layer due to the difference in between the actual output and target output. The outputs from the input layer are propagated to the hidden layer which further propagated to the output layer. The results are propagated back from the output layer to the input layer according to which the weights are updated.



**Fig 2: Artificial Neural Network**

### 4. Experimental results

There are ten parameters which are used in our project they are sex, age, BP, smoking, etc, with these we calculate the risk factor of patient. Table 1 contain the different patient details

**Table No. 1. Details of different patient**

	<b>Patient No. 1</b>	<b>Patient No. 2</b>	<b>Patient No. 3</b>	<b>Patient No. 4</b>	<b>Patient No. 5</b>	<b>Patient No. 6</b>
<b>Sex</b>	Male	Female	Female	Male	Female	Female
<b>Age</b>	50	50	30	35	35	35
<b>Systolic BP</b>	155	155	106	129	129	129
<b>Diastolic BP</b>	96	96	67	75	75	75
<b>Smoking</b>	Yes	No	Yes	Yes	Yes	Yes
<b>Total Cholesterol</b>	6.1	6.1	6.1	10.2	10.2	10.2
<b>HDL Cholesterol</b>	1.7	1.7	0.7	0.3	0.3	0.3
<b>Diabetes</b>	Yes	Yes	No	Yes	No	Yes
<b>LVH</b>	Yes	No	No	Yes	No	No
<b>No. of Years having Diabetes</b>	9	9	10	10	10	10
<b>Risk</b>	36.3	9.8	1.0	64.8	25.3	37.5

## 5. Conclusion

In this paper, a useful and accurate technique for the heart disease prediction is proposed and developed. In this technique we use 2 algorithms; first algorithm is Genetic Algorithm which uses the phenomena of mutation and crossover over various generations. The weights which are used for BP can be optimized. Second algorithm is BPNN, in this optimized output given as input to our network to give much better results. The Neural Network is trained with different topology sizes and no. of iterations to improve the performance of the system.

## References

- Jiang, L., Zhang, H., Cai, Z., & Su, J. (2005). One dependence augmented naive bayes. In *Advanced Data Mining and Applications* (pp. 186-194). Springer Berlin Heidelberg.
- Jabbar, M. A., Deekshatulu, B. L., & Chandra, P. (2012). An Evolutionary algorithm for Heart Disease Prediction. In *Wireless Networks and Computational Intelligence* (pp. 378-389). Springer Berlin Heidelberg.
- Jabbar, M. A., Chandra, P., & Deekshatulu, B. L. (2012, November). Prediction of risk score for heart disease using associative classification and hybrid feature subset selection. In *Intelligent Systems Design and Applications (ISDA), 2012 12th International Conference on* (pp. 628-634). IEEE.
- Jabbar, M. A. (2012). Knowledge discovery from mining association rules for Heart disease Prediction. *Journal of Theoretical and Applied Information Technology*, 41(2), 166-174.
- Chaurasia, V., & Pal, S. (2014). Data Mining Approach to Detect Heart Diseases. *International Journal of Advanced Computer Science and Information Technology (IJACSIT) Vol, 2*, 56-66.

Shouman, M., Turner, T., & Stocker, R. (2012). Applying k-nearest neighbour in diagnosing heart disease patients. *International Journal of Information and Education Technology*, 2(3), 220.

Vijayarani, S. (2013). An Efficient Classification Tree Technique for Heart Disease Prediction. In *International Conference on Research Trends in Computer Technologies (ICRTCT-2013) Proceedings published in International Journal of Computer Applications (IJCA)(0975–8887)* (pp. 6-9).

Anbarasi, M., Anupriya, E., & Iyengar, N. C. S. N. (2010). Enhanced prediction of heart disease with feature subset selection using genetic algorithm. *International Journal of Engineering Science and Technology*, 2(10), 5370-5376.