International Journal of Advanced Technology in Engineering and Science
Vol. No.3, Special Issue No. 01, August 2015
www.ijates.com
ijates
ISSN 2348 - 7550

ANALYSIS OF HEART ATTACK PREDICTION SYSTEM USING GENETIC ALGORITHM

Beant Kaur [1], Dr. Williamjeet Singh [2]

[1]Research Scholar, [2]Asst. Professor, Dept of Computer Engineering, Punjabi University, Patiala (India)

ABSTRACT

Most nations face high and expanding rates of heart diseases or Cardiovascular Disease. Heart disease is a term that assigns to a large number of medical conditions related to heart. These medical conditions describe the abnormal health conditions that directly influence the heart and all its parts. Heart disease is a major health problem in today's time. Heart diseases are mainly because of family background, smoking, high blood pressure etc. This paper aims to implement the data mining technique genetic algorithm for heart disease prediction. The observations reveal that genetic algorithm with 14 attributes give good result. Conclusion from this implementation shows that genetic algorithm has accuracy of 73.46% and takes elapsed time and energy 0.04172s and 1285.99joules respectively. Data ining tool used for the implementation is matlab2013. We note the genetic algorithm parameters accuracy, elapsed time and energy consumption.

Keywords: Data Mining, Genetic Algorithm, Heart Disease, Matlab2013.

I. INTRODUCTION

Due to a wide availability of huge amount of data and a need to convert this available huge amount of data to useful information necessitates the use of data mining techniques. Data Mining and KDD (knowledge discovery in database) have become popular in recent years. The popularity of data mining and KDD (knowledge discovery in database) shouldn't be a surprise since the size of the data collections that are available are far too large to be examined manually and even the methods for automatic data analysis based on classical statistics and machine learning often face problems when processing large, dynamic data collections consisting of complex objects. In [1] a comparative analysis of applications of data mining techniques has been presented.

1.1 Heart Disease

A major challenge facing healthcare organizations (hospitals, medical centers) is the provision of quality services at affordable costs. There are number of factors which increases risk of Heart disease.

I. **Family history of heart disease**
II. **Smoking:**
III. **Cholesterol**
IV. **High blood pressure**
V. **Obesity**
VI. **Lack of physical exercise**

VII. **Poor diet**

VIII. **Physical inactivity**

II. NEED OF STUDY

1. The World Health Organization (WHO) has estimated that 12 million deaths occur worldwide, every year due to the Heart diseases.
2. About 25% deaths in the age group of 25-69 year occur because of heart diseases. In urban areas, 32.8% deaths occur because of heart ailments, while this percentage in rural areas is 22.9.
3. Over 80% of deaths in world are because of Heart disease. WHO estimated by 2030, almost 23.6 million people will die due to Heart disease.
4. The diagnosis of diseases is a significant and tedious task in medicine.
5. Treatment of the said disease is quite high and not affordable by most of the patients particularly in India

III. OBJECTIVES AND RESEARCH GAPS

Objectives of this paper are given below:-

I. To predict the heart attack disease.

II. It helps in reducing treatment costs by providing effective treatments.

III. To find the parameters values in prediction like accuracy, elapsed time and energy consumption.

IV. SURVEY

This paper aims at analyzing the various data mining techniques introduced in recent years for heart disease prediction. [2] Proposed method achieved accuracy values 84.24% and 86.8% for Pima Indians diabetes dataset and Cleveland heart disease dataset, respectively. (Table 1) shows different data mining techniques used in the diagnosis of Heart disease over different Heart disease datasets. Extended the algorithms to deal with non-uniform misclassification costs in order to perform ROC analysis and control the trade-off between sensitivity and specificity**.** [3] Paper presents selected data mining techniques that can be applied in medicine, and in particular some machine learning techniques including the mechanisms that make them better suited for the analysis of medical databases.

Table 1:- Table shows different data mining techniques used in the diagnosis of Heart disease over different Heart disease datasets.

Author	Technique Used	Objective
Carlos, et al.(2001)	association rules	Predict HD
K. Usha Rani(2011)	Classification Neural Networks	Analysis of HD dataset
Jesmin Nahar , et al.(2013)	Apriori Predictive Apriori Tertius	detect factors which contribute to HD in males and females
Latha, et al.(2008)	genetic algorithm CANFIS	IHDP system
M. Akhil jabbar, et al.(2011)	*Clustering* *Association rule mining,* *Sequence number,* Neural Network Naive Bayes	HD prediction
Nan-Chen, et al. (2012)	(EVAR) Machine learning Markov blanket	Intelligent Postoperative Morbidity Prediction of HD
Oleg, et al.(2012)	artificial neural network genetic polymorphisms	Coronary HD diagnosis
Shadab, et al.(2012)	Naive bayes	Prediction system for HD
Shantakumar, et al. (2009)	MAFIA Clustering K-Means	Extraction of Significant Patterns from HD Warehouses for Heart Attack Prediction

[10] Gives the accuracy of 89.01% using neural networks for implementation. [1] Used three data mining techniques J48 classifier, Naive Bayes and Neural Network which gave accuracies 95.56%, 92.42% and 94.85% respectively. [12] Presented prototype model for the breast cancer as well as heart disease prediction using data mining techniques. Two decision tree algorithms C4.5 and the C5.0 have been used on these datasets for prediction and performance of both algorithms is compared. Paper [13] extended the algorithms to deal with non-uniform misclassification costs in order to perform ROC analysis and control the trade-off between sensitivity and specificity. [14] Proposed a system that uses neural network for prediction of heart disease, blood pressure and sugar. A set of 78 records with 13 attributes are used for training and testing. In [15] accuracy of neural networks is 89.01% shown. In paper [16] accuracy of neural networks in prediction of heart disease in weka tool noted approx 100%.

V. TOOL USED FOR IMPLEMENTATION

MATLAB: - MATLAB is a high language and interactive environment for numerical computation, visualization and programming. Using MATLAB we can analyze data, develop algorithms and create models and applications. The language, tool and built-in math functions enable us to explore multiple approaches and reach a solution faster than with spreadsheets of traditional programming languages, such as C/C++ of JAVA.

5.1 Data Set

We take the data set in this paper with 303 records and 14 attributes collected from the online dataset repository of archive.ics.edu/ml/datasets. The dataset parameters are listed as per following:

Table 2:- The Dataset Attributes

Only	14	Attributes	Full description
1	#3	(age)	Patient Age
2	#4	(sex)	Male/Female
3	#9	(cp)	Chest pain type
4	#10	(trestbps)	Resting blood pressure (in mm Hg on admission to the hospital)
5	#12	(chol)	Serum cholestoral (mg/dl)
6	#16	(fbs)	Fasting blood sugar
7	#19	(restecg)	Resting ECG results
8	#32	(thalach)	Maximum heart rate achieved
9	#38	(exang)	Exercise induced angina
10	#40	(oldpeak)	ST depression included by exercise relative to rest
11	#41	(slope)	Slope of the peak exercise ST segment
12	#44	(ca)	Number of major vessels (0-3) colored by fluoroscopy
13	#51	(thal)	3 = normal, 6 = fixed defect, 7 = reversible defect.
14	#58	(num)	Angiographic disease status (Diagnosis of heart disease)

5.2 Algorithms Used for Implementation

Genetic algorithm: - In the computer science field of artificial intelligence, a Genetic Algorithm (GA) is a search heuristic that mimics the process of natural evolution. This heuristic is routinely used to generate useful solutions to optimization and search problems. Genetic Algorithms belong 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. Genetic Algorithms (GAs) are adaptive heuristic search algorithm based on the evolutionary ideas of natural selection and genetics.

VI. ALGORITHMS

Algorithm: - GA for heart disease optimization

STEP 1: Load the heart patient dataset

STEP 2: Validate the dataset in order to make it in the input format for the GA based Heart disease prediction algorithm.

STEP 3: GA finds the optimal solution for the given set of values an returns the personal best and global best values in the form of pBest and gBest, respectively.

$$m_H(i+1) = F_H(i) m_H(i) \left[1 - p_C \frac{l_H}{l-1}\right] [(1-p_m)^{o_H}], \quad (1)$$

STEP 4: Calculate the pBest and gBest on each component of the dataset divided according to the dataset size and the block size.

STEP 5: Then apply fitness function on the basis of rule using Z statistics Z=S(X)-Minimum support/SQRT (min sup*(1-minsup))/N Where S(X) is support of pattern and min sup is user defined threshold

$$f(\boldsymbol{x}) = \sum_{i=1}^{n-1} \left[100 \left(x_{i+1} - x_i^2\right)^2 + (x_i - 1)^2\right] \quad (2)$$

STEP 6: Prune the rules based on Z statistics. After rule evaluation the rules having highest fitness are stored in a pool. Then apply genetic functions on these rules.

STEP 7: Perform single point cross over. Judgment nodes are selected for crossover.

$$GA1: \quad C_{ij} = \Delta C_j \sum_{k=1}^{4} 2^{4-k} a_{4d(i-1)+4(j-1)+k}, \qquad \forall i = \overline{1,n}, \quad j = \overline{1,d}, \quad (3)$$

$$GA2: \quad C'_{ij} = C_{ij} + a_{d(i-1)+j} \Delta C_j, \qquad \forall i = \overline{1,n}, \quad j = \overline{1,d}. \quad (4)$$

STEP 8: Perform mutation by mutating the value of judgment node. This process will be repeated till last generation reached.

STEP 9: After performing GA on each index, calculate gBest, which interprets the best optimal solution given by GA

STEP 10: Build classifiers using the generated Rules on the basis of gBest

STEP 11: Predict the rules on test data using a customized predictivity algorithm.

VII. RESULTS

As the goal of this study is to detect heart disease using data mining technique a classification technique was adopted to develop a predictive model. The implementation was built with three different supervised machine learning algorithm i.e. genetic algorithm using matlab2013. Accuracy graph by using the genetic algorithm is given below. The results have been obtained in the form of the various performance parameters as energy consumption, execution time and accuracy percentage.

The genetic algorithm has been used for the purpose of optimization of the health record data of heart health monitoring. The proposed model has been tested with a standard database in order to obtain the desired results. The results of the parameters have been obtained the graphical form. The graphs by using genetic algorithm for heart disease prediction algorithm have shown below:

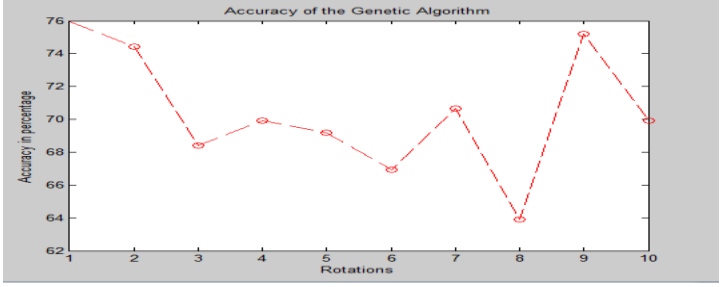

Figure 1:- The Accuracy Graph Obtained with GA Based Heart Disease Prediction Algorithm

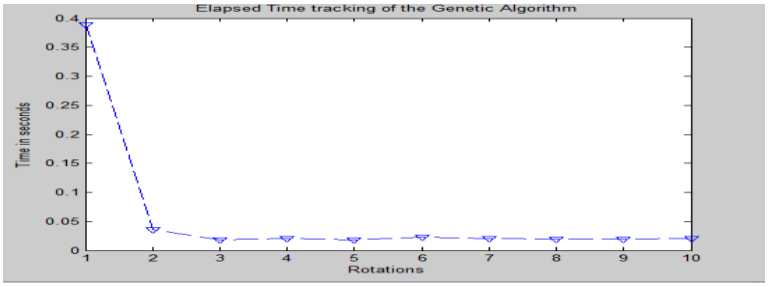

Figure 2:- The Elapsed Time Graph Obtained with GA Based Heart Disease Prediction Algorithm

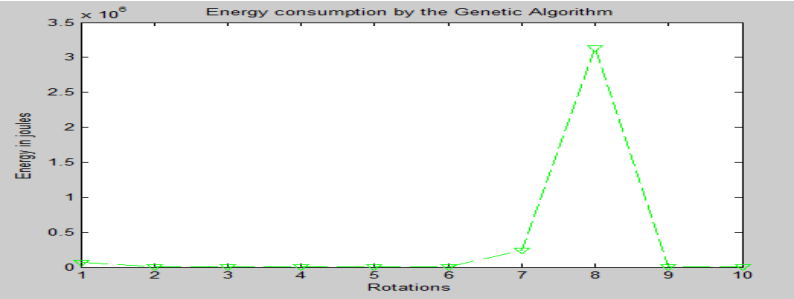

Figure 3:- The Energy Consumption Graph Obtained with GA Based Heart Disease Prediction Algorithm

The energy consumption by genetic algorithm is shown in (Figure 3). The GA based heart disease prediction model is efficient in terms of accuracy (Figure 1). The GA based model consumed almost 125 joules on the average of energy consumption. The total energy consumption on all of the rotations has been almost recorded at approx. 1250 joules. The elapsed time of the GA based heart prediction model (Figure 2) has been recorded 1.72 seconds. The (table 3) is showing the performance parameters of Genetic algorithm based heart prediction model.

Table 3:- The Performance of GA Based Heart Disease Classifier

Average Accuracy for heart disease prediction model using GA	73.46 %
Average Elapsed Time for heart disease prediction model using GA	0.23 seconds
Average Energy Consumption for heart disease prediction model using GA	125.99 joules
Total Elapsed Time for heart disease prediction model using GA	1.72 seconds
Total Energy Consumption for heart disease prediction model using GA	1250.87 oules

VIII. CONCLUSION

In this study, our aim was to implement a data mining technique for prediction of heart diseases and to perform the analysis on the results obtained for different parameters. This paper has provided details on the values of classifiers for the detection of heart disease. Accuracy of genetic algorithm is 73.46%. Energy consumption by genetic algorithm is 1285.99joules and Elapsed time taken in genetic algorithm in overall iterations is 0.0340s.

IX. FUTURE SCOPE

As future work, we will process the new data mining techniques like Classification and Regression Trees algorithm with other new algorithm to refine result more accurately. We will work to predict more accurately the presence of heart disease with reduced number of attributes. We have planned to design and develop an efficient heart attack prediction system with the aid of these selected significant patterns using artificial intelligence techniques.

REFERENCES

[1]. S.A. Lashari, et al.," Comparative Analysis of Data Mining Techniques for Medical Data Classification", Proceedings of the 4th International Conference on Computing and Informatics, ICOCI, 28-30 August, 2013 Sarawak, Malaysia. University Utara Malaysia, Paper No.-034, 2013.

[2]. Humar Kahramanli, et al. "Design of a Hybrid System for The Diabetes and Heart Diseases" Elsevier, Expert Systems with Applications, Vol-35, pg no.82–89, 2008.

[3]. Nada Lavrac, "Selected techniques for Data Mining in Medicine", Elsevier, Artificial Intelligence in Medicine, Vol-16, pg no.3–23, 1999

[4]. Carlos Ordonez, et al., "Mining Constrained Association Rules to Predict Heart Disease", IEEE. Published in International Conference on Data Mining, pg no.433-440, 2001.

[5]. K. Usha Rani, "Analysis of heart diseases dataset using neural network approach", International Journal of Data Mining & Knowledge Management Process, Vol.-1, Issue-5, pg no.1-8, 2011.

[6]. Jesmin Nahar, et al., "Association rule Mining to Detect Factors which Contribute to Heart Disease in Males and Females", Elsevier, Expert Systems with Applications, Vol-40, pg no.1086–1093, 2013.

[7]. M. Akhil. Jabbar, et al., "Cluster Based Association Rule Mining for Heart Attack Prediction", Journal of Theoretical and Applied Information Technology, Vol.-32, Issue-2, pg no.196-201, 2011.

[8]. Nan-Chen Hsieh, et al., "Intelligent Postoperative Morbidity Prediction of Heart Disease Using Artificial Intelligence Techniques", Springer, J Med Syst, Vol-36, pg no.1809–1820, 2012.

[9]. Oleg Yu. Atkov, et al., "Coronary Heart Disease Diagnosis by Artificial Neural Networks including Genetic Polymorphisms and Clinical Parameters", Elsevier, Journal of Cardiology, Vol-59, pg no.190-194, 2012.

[10]. Resul Das, et al., "Diagnosis of Valvular Heart Disease through Neural Networks Ensembles", Elsevier, Computer Methods and Programs in Biomedicine, Vol-93, page no.185–191, 2009.

[11]. Abhishek Taneja, "Heart Disease Prediction System using Data Mining Techniques", Oriental Scientific Publishing Corporation India, Vol.-6, Issue-4, Pg no. 457-466, 2013.

[12]. Mohammad Taha Khan, et al., "A Prototype of Cancer/Heart Disease Prediction Model Using Data Mining", International Journal of Applied Engineering Research, Vol-7, Issue-11, 2012.

[13]. Matjaz Kukar, et al., "Analyzing and improving the diagnosis of Ischaemic Heart Disease with Machine Learning ", Elsevier, Artificial Intelligence in Medicine, Vol-16, pg no.25–50, 1999

[14]. Niti Guru, et al.," Decision Support System for Heart Disease Diagnosis Using Neural Network", Delhi Business Review, Vol. 8, Issue-1, 2007.

[15]. Resul Das, et al., "Effective diagnosis of Heart Disease through Neural Networks Ensembles", Elsevier, Expert Systems with Applications, Vol-36, pg no.7675–7680, 2009.

[16]. Chaitrali S. Dangare, et al., "A Data Mining Approach for Prediction of Heart Disease using Neural Networks", International Journal of Computer Engineering and Technology, Vol-3, Issue-3, pg no. 30-40, 2012.