

**INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH
TECHNOLOGY****A REVIEW ON THE DETECTION OF HEART ATTACK USING DATA MINING
BY ACO TECHNIQUE****Pise Satish Prakashrao^{*1}, Anoop Singh² & Ritesh Kumar Yadav³**^{*1}Research Scholar, Department of Computer Science & Engineering, RKDFIST, Bhopal, India²Asst. Professors, Department of Computer Science & Engineering, RKDFIST, Bhopal, India³Asst. Professors, Department of Computer Science & Engineering, RKDFIST, Bhopal, India

DOI: 10.5281/zenodo.1165712

ABSTRACT

The goal of data mining is to extract knowledge from huge amount of data. Now a day's data mining technique used in the field of medical diagnose of critical diesis and clinical data. In this research propose model give a solution to predict heart diseases. In this paper proposes a novel approach of applying the Ant Colony Optimization technique (ACO) for extracting the Association Rules (AR) from the database to detect heart attack. This algorithm is broadly are many types of heart disease which are considered here Congenital Heart Disease Congestive Heart Failure Coronary Heart Disease. According to risk level identified we find the max pheromone value, max pheromone value is the addition of weight and the risk level. In the early days many research has been done in the field of considerable diseases like Heart Attack using various technologies like soft computing, Fuzzy Techniques and Data mining techniques. This study leads to make such kind of diseases efficiently identifiable and curable. So by this approach we can increase the detection probability in the early stage which is not generally detected in the earlier stage.

KEYWORDS: Heart Disease, Ant Colony Optimization, Spectrums K-Means Algorithm, pheromone, etc.**I. INTRODUCTION**

Data mining technique means the use of sophisticated data analysis tools to determine previously unknown, valid patterns and relationships in large data set. These tools can include statistical models, mathematical algorithm and machine learning methods in early detection of chronic disease. Medical data mining has great potential for exploring the hidden patterns in the data sets of the medical domain. These patterns can be utilized for clinical diagnosis. However, the available raw medical data are widely distributed, heterogeneous in nature, and voluminous. These data need to be collected in an organized form. This collected data can be then integrated to form a hospital information system. Data mining technology provides a user oriented approach to novel and hidden patterns in the data [1]. Heart disease has considerably increased for the last two decades and become the leading cause of death for people in most of the countries in the world. World Health Organization (WHO) reported that 30% of death is due to heart disease [1&2]. In 2008 17.3 million people died due to heart disease. More than 80% of passing away in world is because of coronary illness.

According to survey of WHO, 17 million total global deaths are due to heart attacks and strokes. The deaths due to heart disease in many countries occur due to work overload, mental stress and many other problems. On the whole it is found as primary reason behind death in adults. Diagnosis is complicated and important task that needs to be executed accurately and efficiently. The diagnosis is often made, based on doctor's experience & knowledge. This leads to unwanted results & excessive medical costs of treatments provided to patients [15-17].

II. SYSTEM DESCRIPTION

Data Mining involves few steps as in figure 1 from raw data collection to some form of new knowledge. The iterative process consists of the following steps like Data Cleaning, Data Integration, Data Selection, Data Transformation, Data Mining, Pattern Evaluation and Knowledge Representation. Hence conceptually it is discovery of knowledge from datasets. Data mining can quickly answer business questions that would have

[Prakashrao * *et al.*, 7(2): February, 2018]
ICTM Value: 3.00

otherwise consumed a lot of time. Applications of Data mining spread in different fields Some of like Market Behaviour contains identification of customer activities for analysis [3-6]. It performs various functionalities to mine the data from dataset as given below:

- Pre-processing: It is the process of scattering out anomalies and useless unimportant data which is going to be mined. in this some steps taken like adding missing values, generalizing, aggregating, compressing and normalizing.
- Clustering: It is partitioning a large set of data into related sub-classes.
- Classification: It is tagging or classifying data items into different user-defined categories.
- Associative analysis: It helps in bringing out hidden relationships among data items in a large data set.
- Regression: It is used to predict values of a dependent variable by constructing a model or a mathematical function out of independent variables.
- Summarisation: It helps in coming up with a compact description for the whole data set.

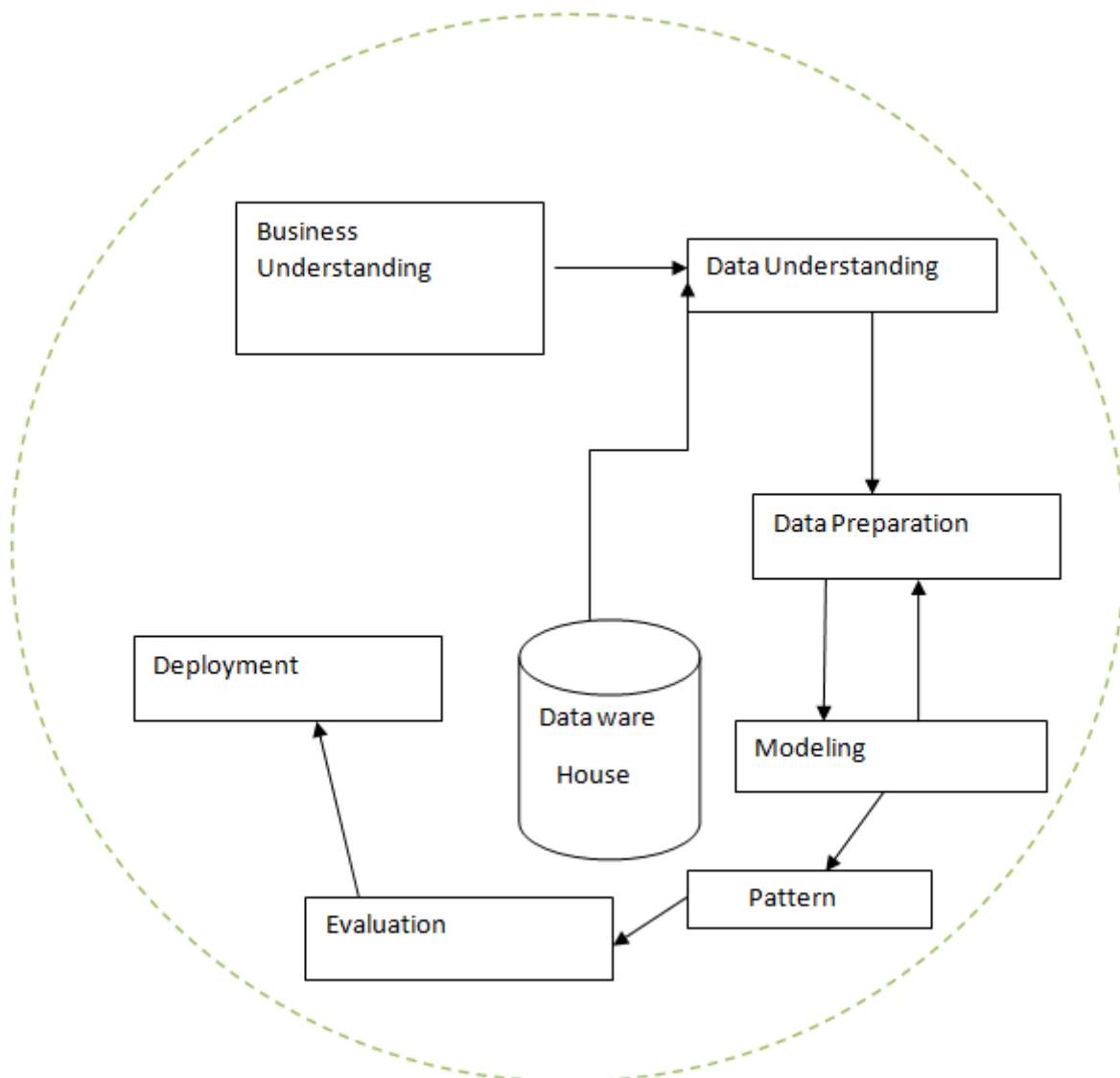


Figure : 1 Data Mining Phases

Heart Disease

The rise of health care cost is one of the world's most important problems [17]. Heart attack happens when there is irregularity in the flow of blood and heart muscle is injured because of inadequate oxygen supply. World Health Organization in the year 2008 reported that 30% of total global deaths are due to Cardio Vascular Disease (CVD). By 2030, almost 25 million people will die from CVDs, mainly from heart disease and stroke. To manage the data of patients regarding their healthcare hospitals widely uses information storage systems by which large amount of data generates. Now the challenge is to make this information useful. There is prosperously hidden exists in such storages. So we need to extract correct and feasible data from here so that correct clinical decisions can be taken. Hence the main objective of this research is to develop a Support based Ant Colony Optimization Framework for Heart Disease Prediction.

III. RELATED WORK

In 2014 Dubey A. et al [4] India is set to witness a spike in deaths due to heart diseases. Early stage detection may prevent the death due to the heart diseases. In this paper we provide an efficient approach which is based on Data Mining and Ant Colony Optimization technique (DMACO) for Heart Disease Prediction so that we can prevent it in the earlier stages. For this we first taken the concept of data mining to finding the support, generated support is used as a weight of the symptom which will be the initial pheromone value of the ant. Then we consider Pain in the chest, Discomfort radiating to the back, choking feeling (heartburn), Nausea, Extreme weakness and Irregular heartbeats as the factor of heart attack. According to risk level identified we find the max pheromone value), max pheromone value is the addition of weight and the risk level. After applying the DMACO algorithm we can observe the increasing detection rate. So by this approach we can increase the detection probability in the early stage which is not generally detected in the earlier stage.

In 2014 Dr. Durairaj.M, Sivagowry.S el at[5] Medical Ecosystem is originated with rich information database, but inadequate in techniques to extract the information from the database. This is because of the lack of effective analysis tool to discover hidden relationships and trends in them. By applying the data mining techniques, valuable knowledge can be extracted from the health care system. Extracted knowledge can be applied for the accurate diagnosis of disease and proper treatment. Heart disease is the leading cause of death in all over the world over past ten years. Researchers have developed many hybrid data mining techniques for diagnosing heart disease. Here a preprocessing technique and analysis of the accuracy for prediction after preprocessing the noisy data explained. It is also observed that the accuracy has been increased to 91% after preprocessing. Swarm Intelligence techniques hybridized with Rough Set Algorithm are to be taken as future work for exact reduction of relevant features for prediction

In 2014 Masethe H.D. el at [6] The heart disease accounts to be the leading cause of death worldwide. It is being found tough to predict the heart attack as it is a complex task that requires experience and knowledge to medical experts. The health sector today contains hidden information that can be important in making decisions. Here some mining algorithms like Naïve Bayes, REPTREE, J48, CART, and Bayes Net are used for the efficient prediction heart attacks. The research result found prediction accuracy of 99%.

In 2013 Kumar S.,Kuar G el at[7] the Nowadays the use of computer technology in the fields of medicine area diagnosis, treatment of illnesses and patient pursuit has highly increased The objective of this paper is to detect the heart diseases in the person by using Fuzzy Expert System. The designed system based on the Parvati Devi hospital, Ranjit Avenue and EMC hospital Amritsar and International

Lab data base. The system consists of 6 input fields and two output field. Input fields are chest pain type, cholesterol, maximum heart rate, blood pressure, blood sugar, old peak. By the obtained result field presence of heart disease in the patient and precautions accordingly has been detected. It is integer valued from 0 (no presence) to 1 (distinguish presence (values 0.1 to 1.0). We can use the Mamdani inference method. The generated outcomes developed system are comparatively analyzed. This observation found correct 92%.

In 2012, Muhammed et al. [8] present and discuss the experiment that was executed with naïve bayes technique in order to build predictive model as an artificial diagnose for heart disease based on data set which contains set of parameters that were measured for individuals previously. Then they compare the results with other techniques according to using the same data that were given from UCI repository data.

In 2012 Dangare c.s. et al [9] present in the Healthcare industry is generally “information rich”, but unfortunately not all the data are mined which is required for discovering hidden patterns & effective decision making. To discover knowledge from database and for the purpose of medical research, particularly in Heart disease prediction involvement and use of specific techniques of data mining required and implemented. This paper has analyzed prediction systems for Heart disease using more number of input attributes. The system uses medical terms such as sex, blood pressure, cholesterol like 13 attributes to predict the likelihood of patient getting a Heart disease. Until now, 13 attributes are used for prediction. This research paper added two more attributes i.e. obesity and smoking. With the different classification methods Heart Attack databases analyzed. The performance of these techniques is compared, based on accuracy. As per our results accuracy of Neural Networks, Decision Trees, and Naive Bayes are 100%, 99.62%, and 90.74% respectively. By the analysis it is found that Neural Network technique predicts Heart disease with highest accuracy.

In 2012 bhatla N. el at[10] present and discuss Cardiovascular disease is a term used to describe a variety of heart diseases, illnesses, and events that impact the heart and circulatory system. A medical experts use several sources of data and tests to make a diagnostic impression but it is not necessary that all the tests are useful for the diagnosis of a heart disease. The objective of our work is to reduce the number of attributes used in heart disease diagnosis that will automatically reduce the number of tests which are required to be taken by a patient. Our work also aims at increasing the efficiency of the proposed system. The observations illustrated that Decision Tree and Naive Bayes using fuzzy logic has outplayed over other data mining techniques.

In 2008, Tsipouras, Markos G. et al. [11] presented a fuzzy rule-based decision support system (DSS) for the diagnosis of coronary artery disease (CAD). The system is automatically generated from an initial annotated dataset, using a four stage methodology. The dataset used for the DSS generation and evaluation consists of 199 subjects, each one characterized by 19 features, including demographic and history data, as well as laboratory examinations. Here a cross validation is applied 10 times which produces average sensitivity 62% and specificity 54%, using the set of rules extracted from the decision tree (first and second stages), while the average sensitivity and specificity increase to 80% and 65%, respectively, when the fuzzification and optimization stages are used. The system provides CAD diagnosis based on easily and noninvasively acquired features, and is able to provide interpretation for the decisions made.

In 2010, Yosawin Kangwanariyakul, Chanin Nantasenamat et al. [12] tells about the Ischemic Heart Disease (IHD) which is a major cause of death. Early and accurate detection of IHD along with rapid diagnosis are important for reducing the mortality rate.

In 2010, Srinivas, K. et al. [13] tell that Heart disease (HD) is a major cause of morbidity and mortality in the modern society. The preparation of medical diagnosis is noteworthy keen observation process but it is important to do so perfectly and accurately with efficiently.

In 2012, Muhammed et al. [14] present and discuss the experiment that was executed with naïve bayes technique in order to build predictive model as an artificial diagnose for heart disease based on data set which contains set of parameters that were measured for individuals previously. Then they compare the results with other techniques according to using the same data that were given from UCI repository data.

A. Various Data Mining Tools & Techniques

Since knowledge set contains raw data set from where it is required to extract feasible data, hence technique involved for the same is data mining comprises with variety of sub techniques and algorithms. Some of the techniques and tools (Table 1) are collectively studied and mentioned here.

- ACO: The Ant Colony Optimization (ACO) algorithm is a meta heuristic that has a sequence of distributed computation, autocatalysis (positive feedback), and constructive greediness to find an optimal solution for combinatorial optimization problems. This algorithm tries to mimic the ant's behaviour in the real world.
- Knn Classifier : Nearest neighbour classifiers are based on learning by analogy. The n different training patterns are described by n dimensional numeric attributes. Here every pattern is like a point oc corresponding dimensional space. In this way, all of the training patterns are stored in an n-dimensional pattern space. From which p-nearest neighbour pattern classifier searches the pattern space for the p training patterns that are closest to the unknown sample.

- Closeness: It is defined in terms of Euclidean distance, where the Euclidean distance, where the Euclidean distance between two points, $X=(x_1,x_2,\dots,x_n)$ and $Y=(y_1,y_2,\dots,y_n)$ is $\text{dist}(X_1,X_2)=$ The unknown sample is assigned the most common class among its k nearest neighbours.
- SVM: This Support vector Machine is binary classifier, the performance of classification of support vector machine is high in comparison of another binary classifier such as decision tree, KNN and bayes classifier. It has been successfully applied to numerous classification and pattern recognition problems such as text categorization, image recognition and bioinformatics.
- K Mean Clustering: K means is used to solve clustering problem and is known as unsupervised learning algorithm. The process is very efficient way to classify the data set into number of clusters (assume k clusters). The k -means algorithm takes the input for the number of clusters (say k) and divides a set of n objects into k clusters so that the emerging inter-cluster similarity is low but the intra-cluster similarity is high. According to the mean value of the object present in a cluster, cluster similarity is measured.
 1. Artificial Neural Network: It is widely used machine learning algorithm, which is equivalent to the Neurons in Humans. In ANN, artificial neurons and process information are interconnected using suitable connections for computation. It is a learning system that changes its structure based on information (i.e. external or internal) that progresses through the network in the phase of learning. It processes in two ways: 1. Feed Forward Network 2. Back Propagation Algorithm
- Naïve bayes: Naïve bayes classifier is based on bayes theorem. It uses conditional independence, means it assumes that an attribute value on a given class is independent of the values of the other attributes.
- Decision trees: The decision tree approach is more powerful for classification problems. There are two steps in this techniques building a tree and applying the tree to the dataset. There are many popular decision tree algorithms CART, ID3, C4.5, CHAID, and J48.

Table : 1 Data Mining Tools

S.No.	Tool Name	Platform	Descriptions	Features
1	Rattle	Linux, Mac OS, Windows	Developed using R statistical programming language.	Clustering Modelling visualization
2	KEEL	Based on Java	Knowledge Extraction for Evolutionary Learning	Pre-processing Evolutionary learning
3	Apache Mahout	Hadoop	Mahout is primarily a library of machine learning algorithms	Clustering Classification Pattern Mining
4	ELKI	JAVA	It is open source software written in Java and licensed under AGPLv3	Cluster Analysis Detection Scalability Performance
5	MOA	Distributed under GNU-GPL JAVA API	Massive Online Analysis	Real time applications Stream Mining
6	WEKA	JAVA based	Clustering Filtering Classification Association selection	Filtering Clustering Classification Attribute selection
7.	Rapid Miner	Python R statistical language	It is leading predictive analytic platform	Built-in template Repeatable workflow Professional

				visualization
8	Knime	Written in JAVA on Eclipse	It is one of the leading open source analytic, integration and reporting platforms	Collection Analysis Modelling Pr-processing
9	DMelt	Linux Mac OS Windows Own IDE	It is a computational platform, offering statistics, numeric and symbolic computations, scientific visualisation	Linear regression Curve fitting Cluster analysis Fuzzy algorithm Analytic calculations

B. Risk Analysis

To evaluate the performance and measure of system stability some parameters are calculated and analyzed. Some of them are mentioned below:

- Sensitivity : It is rate of change of result generated by the system on applying different input and represented as Sensitivity = True Positive/ Positive
- Specificity: It is related to classifier's ability to identify negative results. Consider the example of medical test used to identify a certain disease. The specificity of the test is the proportion of patients that do not to have the disease and will successfully test negative for it. It can be represented as :
Specificity = True Negative/ Negative
- Precision: It is a measure of retrieved instances that are relevant. It can be represented as :
Precision = True Positive/(True Positive + False Positive)
- Accuracy: This is the simplest scoring measure. It calculates the proportion of correctly classified instances. It can be represented as :

Accuracy = (True Positive + True Negative) / (Positive + Negative)

- True Positive (TP): If the instance is positive and it is classified as positive
- False Negative (FN): If the instance is positive but it is classified as negative
- True Negative (TN): If the instance is negative and it is classified as negative
- False Positive (FP): If the instance is negative but it is classified as positive

IV. Conclusion

There are large number of systems had been already in use for medical disease symptoms detections using data mining concepts. The use of data mining techniques in Heart Disease Detection increases the chance of making a correct and early detection, which could prove to be vital in combating the disease. We proposed an efficient concept for detection of Heart Diseases based on ant colony optimization and data mining.

V. REFERENCES

- [1]. Hnin Wint, Khaing, "Data Mining based Fragmentation and Prediction of Medical Data", IEEE International Conference on Computer Research and Development, Vol.2 pp.480-485, March 2011.
- [2]. C.Helma, E.Gottmann, and S.Kramer, "Knowledge discovery and data mining in toxicology", Statistical methods in medical research, Vol. 9, pp. 329-358, August 2000.
- [3]. S.J.Hamilton, G.Chew and G. Watts, "Therapeutic regulation of endothelial dysfunction in type 2 diabetes mellitus", Diabetes and Vascular Disease Research Vol.4 No.89, June 2007.
- [4]. Animesh Dubey, Rajendra Patel and Khyati Choure, "An Efficient Data Mining and Ant Colony Optimization technique (DMACO) for Heart Disease Prediction" , International Journal of Advanced Technology and Engineering Exploration, Volume-1 Issue-1 ,pp 1-5 , December-2014.
- [5]. Dr. Durairaj.M, Sivagowry.S, "A Pragmatic Approach of Preprocessing the Data Set for Heart Disease Prediction", International Journal of Innovative Research in Computer and Communication Engineering, Vol. 2, Issue 11, pp- 6457-6465, November 2014

- [6]. Hlaudi Daniel Masethe, Mosima Anna Masethe, "Prediction of Heart Disease using Classification Algorithms", Proceedings of the World Congress on Engineering and Computer Science, Vol 2, pp 22-24 October, 2014,
- [7]. Sanjeev Kumar, Gursimranjeet Kaur, "Detection of Heart Diseases using Fuzzy Logic", International Journal of Engineering Trends and Technology (IJETT) ,Volume 4 Issue 6, pp 2694-2699, June 2013.
- [8]. Muhammed, Lamia AbedNoor, "Using data mining technique to diagnosis heart disease", ICSSBE International Conference on Statistics in Science, Business and Engineering, pp.1-3, September 2012.
- [9]. Chaitrali S. Dangare Sulabha S. Apte, "Improved Study of Heart Disease Prediction System using Data Mining Classification Techniques", International Journal of Computer Applications, Volume 47, pp 44-48, June 2012.
- [10]. Nidhi Bhatla Kiran Jyoti, "A Novel Approach for Heart Disease Diagnosis using Data Mining and Fuzzy Logic", International Journal of Computer Applications, Volume 54, pp 16-21 September 2012.
- [11]. Tspouras, G.Markos and Fotiadis I.D., "Automated Diagnosis of Coronary Artery Disease Based on Data Mining and Fuzzy Modeling" IEEE Transactions on Information Technology In Biomedicine, Vol.12(4), pp. 447-458, July 2008.
- [12]. Kangwanariyakul Y., Chanin N., Tanawut T., Thanakorn N., "Data Mining of Magnetocardiograms for Prediction of Ischemic Heart Disease" EXCLI Journal, Vol.33 (9), pp.82-95, July 2010.
- [13]. Srinivas K., Rao, G.R. Govardhan, "Analysis of coronary heart disease and prediction of heart attack in coal mining regions using data mining techniques", IEEE International Conference on Computer Science and Education, pp. 1344-1349, August 2010.
- [14]. Muhammed, Lamia AbedNoor, "Using data mining technique to diagnosis heart disease", ICSSBE International Conference on Statistics in Science, Business and Engineering, pp.1-3, September 2012.
- [15]. <http://www.cs.columbia.edu/~wfan/software.htm>.
- [16]. <http://people.revoledu.com/kardi/tutorial/kMean/NumericalExample.htm>
- [17]. www.google.com