ISSN: 2277-5528
Impact Factor- 4.015

INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & MANAGEMENT

HEART ATTACK DETECTION BY IMPROVING DATA MINING USING ACO TECHNIQUE

Pradeep Kumar Pandey*and Ruchi Dronawat**
Department of Computer Science Engineering, SIRT, Bhopal(M.P.)
*Pradeep65656@gmail.com, **dronruchi@gmail.com

Abstract

Heart disease is a term that assigns a large number of medical conditions to the heart. These medical conditions describe abnormal health conditions which directly affect the heart and all its parts. There are many types of heart disease that can be considered congenital heart disease here, cardiac failure coronary heart disease. Based on the identified risk, we identify the maximum value of pheromones; The maximum value of pheromone is a combination of weight and risk level. The next step of ant is to find the maximum value of the pheromone because the ratio of the speed of the sensitive ant and its properties will change. With this approach, the number of pieces can be managed through the ACO parameter. In this research, we provide an efficient approach that is based on data mining and ant colony optimization techniques (ACO) for predicting cardiovascular disease so that we can stop it in earlier steps. For this, we first took the concept of data mining to find support; generated support is used as the weight of the symptom, which will be the initial pheromone value of ant. Then we consider pain in the chest, radiation on the back, feeling suffocated (heartburn), nausea, excessive weakness, and irregular heartbeat, a heart attack. This research work uses benchmark performance metrics i.e. sensitivity, specificity, and classification accuracy.

Keywords: ACO, Heart Disease, K-Means Clustering, Naïve Bayes, Pheromone, Spectrums.

Introduction

Many factors in everyday life affect the human heart. Many problems are occurring at a rapid pace and new cardiovascular diseases are being identified rapidly. Today, in the world of stress, being an essential organ in a human body which pumps blood to the body through blood for circulation and preserving its health for a healthy life. The health of the human heart is based on the experiences of a person's life and it is entirely dependent on the individual's professional and personal behaviors. There may also be several genetic factors through which a type of cardiovascular disease goes down from generation.

A heart attack is increasing every day in this modern world. According to the World Health Organization (WHO), an estimated 17 million people die from heart disease every year, especially heart attacks and strokes [1]. Therefore, it is necessary to record the most important symptoms and health habits which contribute to a heart attack. Various tests are conducted before the diagnosis of heart attack and strokes. including auscultation, ECG, cholesterol, blood pressure, and blood sugar. These tests are often long and long when a patient's condition can be serious and he should start taking medication immediately, so it is important to prioritize the tests [2].

Therefore, an automated medical diagnostic system might be very beneficial by bringing them all together. Appropriate computer-based information and / or decision support systems can help in obtaining clinical trials at a lower cost. A comparative study of the various techniques available for efficient and accurate implementation of the automated system is essential. The purpose of this research is to analyze the proposed various forecast / descriptive data mining techniques in recent years to diagnose heart disease.

ISSN: 2277-5528
Impact Factor- 4.015

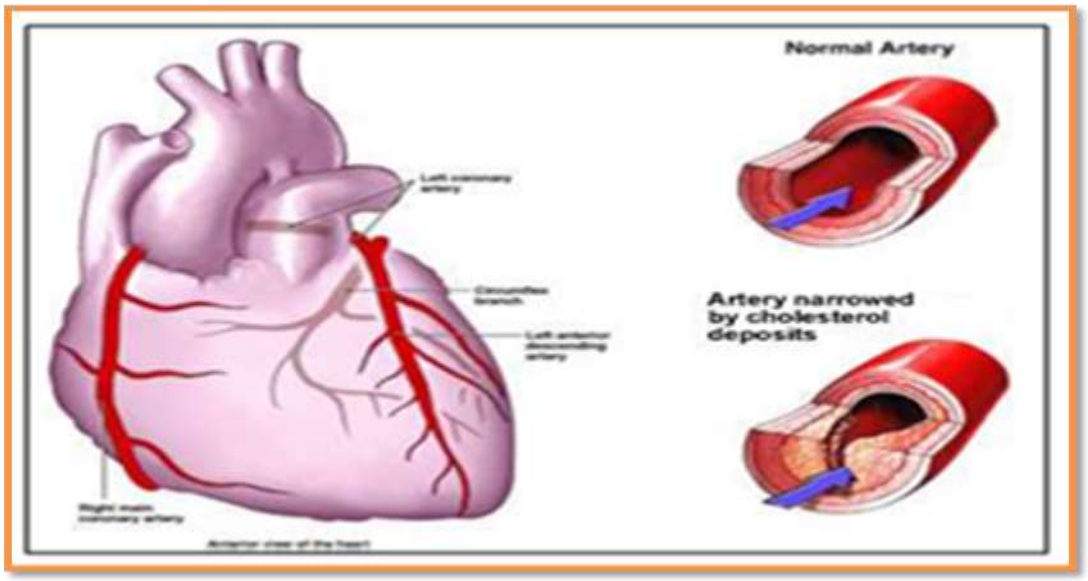

Figure 1: image of heart disease

Proposed Methodology

This section includes our K-Means and ACO based convenience selection and classification system. The main structure of the proposed system is shown in Figure 2. Our system includes convenience choices based on K-Nearest neighbor, support vector machine, Naïve Bayes, random forests and clusters. Training datasets are prepared according to the binary class classification problem.

Datasets are selected from the training, the features are selected, then the best subsets of facilities are optimized by our combined K-Means and ACO algorithms, and then, using the best features selected, new features are classified with data mining software implemented in Java.

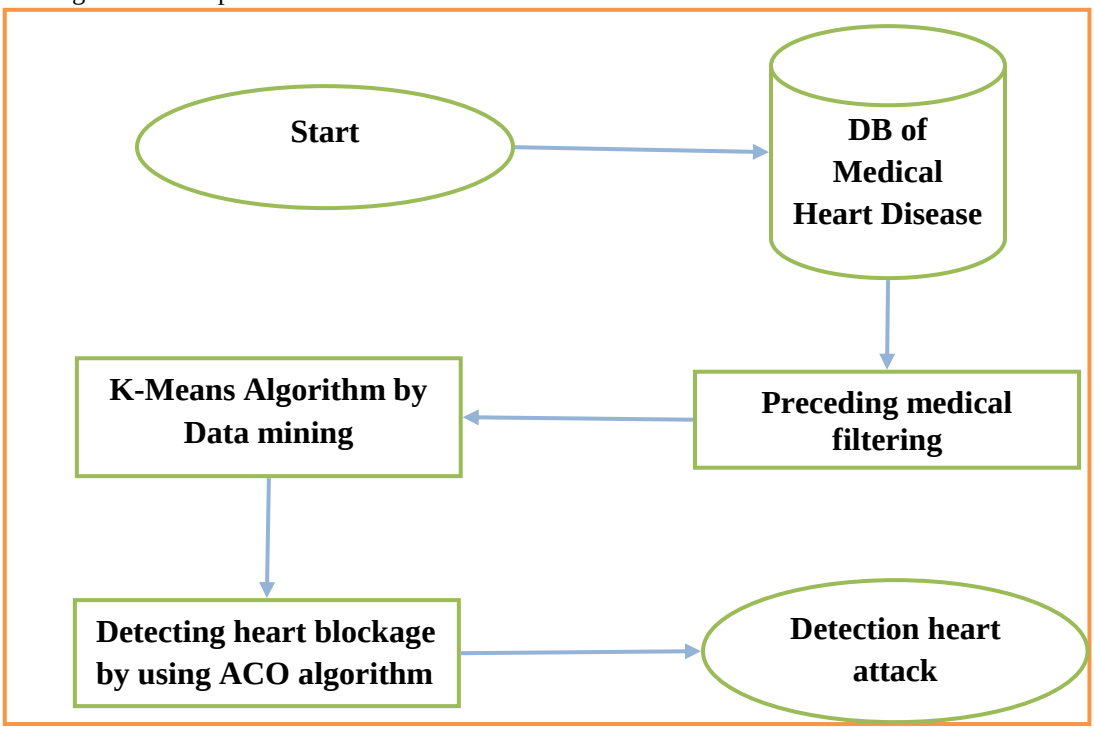

Figure 2: The proposed architecture for heart attack detection

The population of the ants (ants) collectively considers the problem using graph representation. Although each ant is able to find a (perhaps poorly) solution, a mass interaction between the ants can result in a good quality solution. Pheromone trails encode long-term memory about the whole ant-search process. Its value depends on the problem representation and optimization objective.

ACO algorithm

The main purpose of the Ant Colony Optimization Method is to select a subset of features and reduce redundancy among them. In this method, each ant selects the lowest equality characteristics in relation to the previously selected features. Therefore, if an attribute is selected by most ants, then it indicates that the features have the lowest resemblance with other characteristics. The facilities receive the largest amount of pheromone, and the possibility of its selection by other ants will increase in subsequent iterations. Finally, considering the similarity between the features, the selected main features will have more pheromone values. Thus, the ACO method selects the best features with minimal redundancy. Figure 3.4 shows the depiction of the feature selection problem.

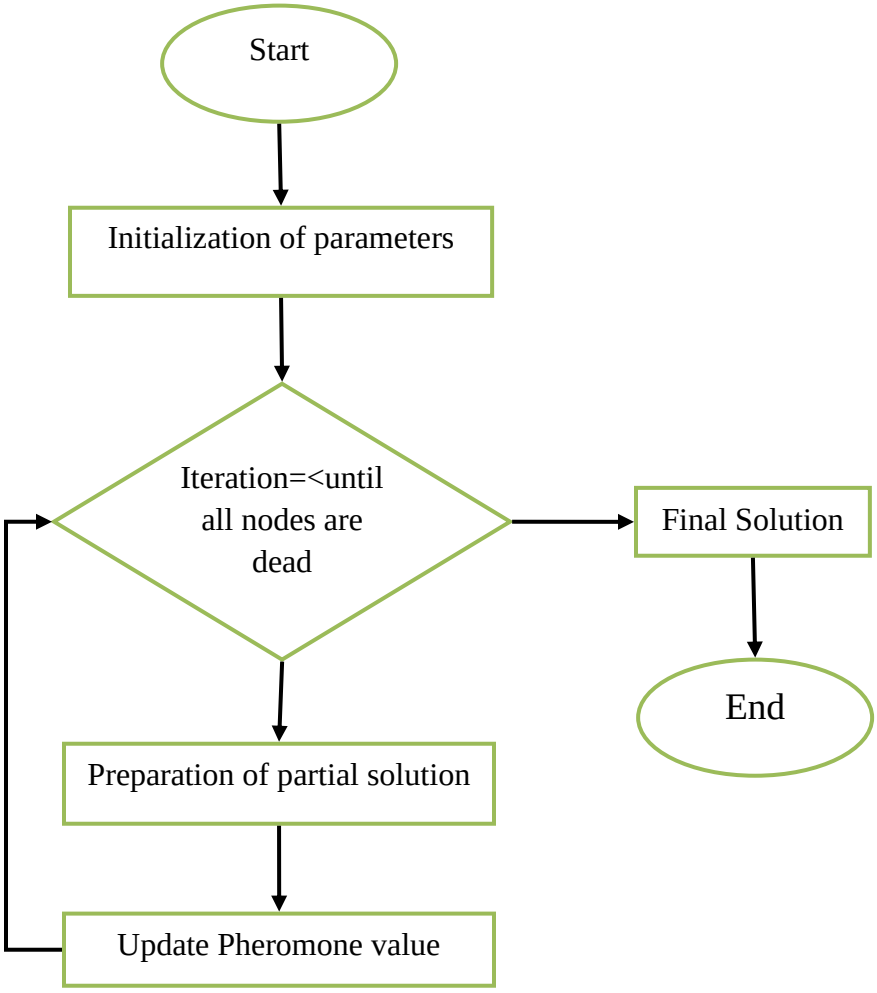

Figure 3: The flowchart of the ACO algorithm

Algorithm 1 : Association Rule Data Mining

C1 ¬ init-pass(T);
F1 ¬ {f | f Î C1, f.count/n ³ minsup}; // n: no. of transactions in T
for (k = 2; Fk-1 ¹ Æ; k++) do
Ck ¬ candidate-gen(Fk-1);
for each transaction t Î T do

```
                    for each candidate c Î Ck do
                        if c is contained in t then
                          c.count++;
                    end
                end
            Fk ¬ {c Î Ck | c.count/n ³ minsup}
        end
return F ¬ Uk Fk;
```

Algorithm 2: ACO algorithm using heart disease detection

Input : DB= Initial training Dataset
T= Test Data
Outputs: E={e1,e2,e3,…,en}entities
[1] get the initial dataset DB as Cluster variables =k_m
[2] set criteria for fitness f = No
[3] check fitness of initial dataset
[4] repeat
[5] Employee ANT Pheromone
[6] Onlooker ANT Pheromone
[7] Scout ANT Pheromone
[8] Calculate fitness measure f
[9] Criteria matching if exists then f= yes
[10] end until f=yes
[11] return E

Result analysis Initial Information

The Initial Information shows two basic operation. If you click on the first lable that is master database , then the next screen will open which shows the corresponding values and their weight.

Result Evaluation

The results of our experimental analysis in finding significant patterns for heart attack prediction are presented in the above section. With the help of the database, the patterns significant to the heart attack prediction are extracted using the approach discussed

The sample combinations of heart attack parameters for normal and risk level along with their values and levels are detailed below. In that, lesser value (0.1) of weight comprises the normal level of prediction and higher values other than 0.1 comprise the higher risk levels.

After applying ACO several time, we receive the value as shown in figure 4. Based on the above grouping we received the final pheromone trails as shown in figure 5(a) & (b)

ISSN: 2277-5528
Impact Factor- 4.015

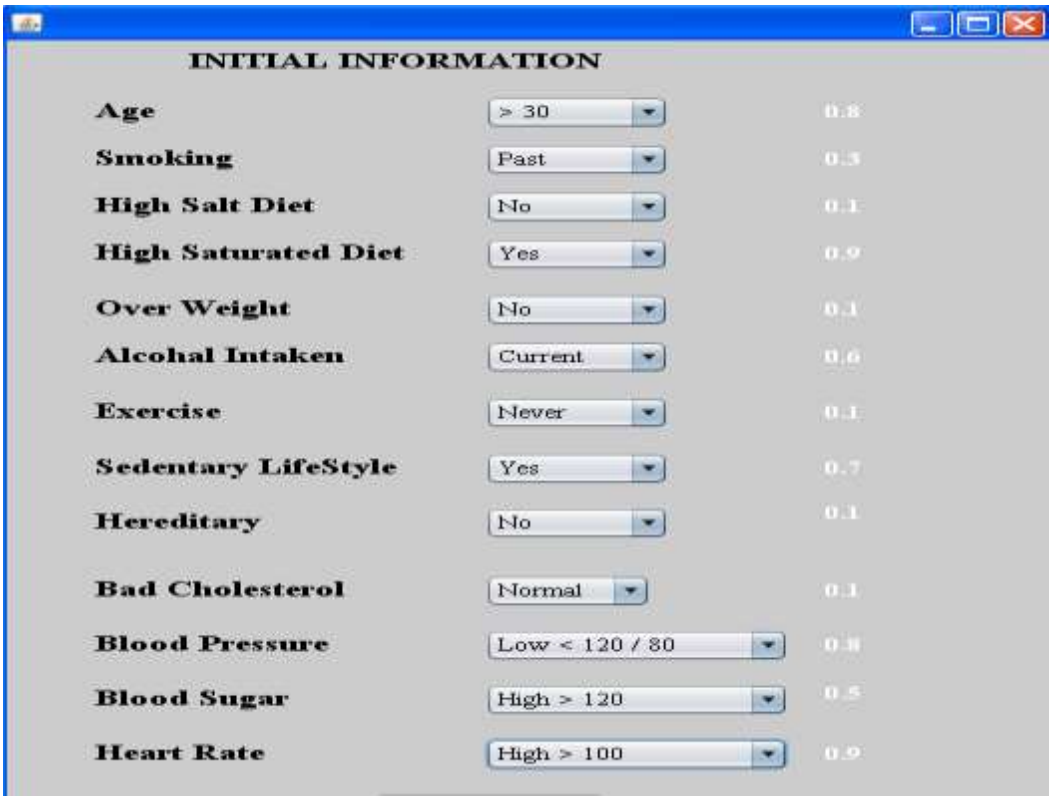

Figure 4: Initial Information

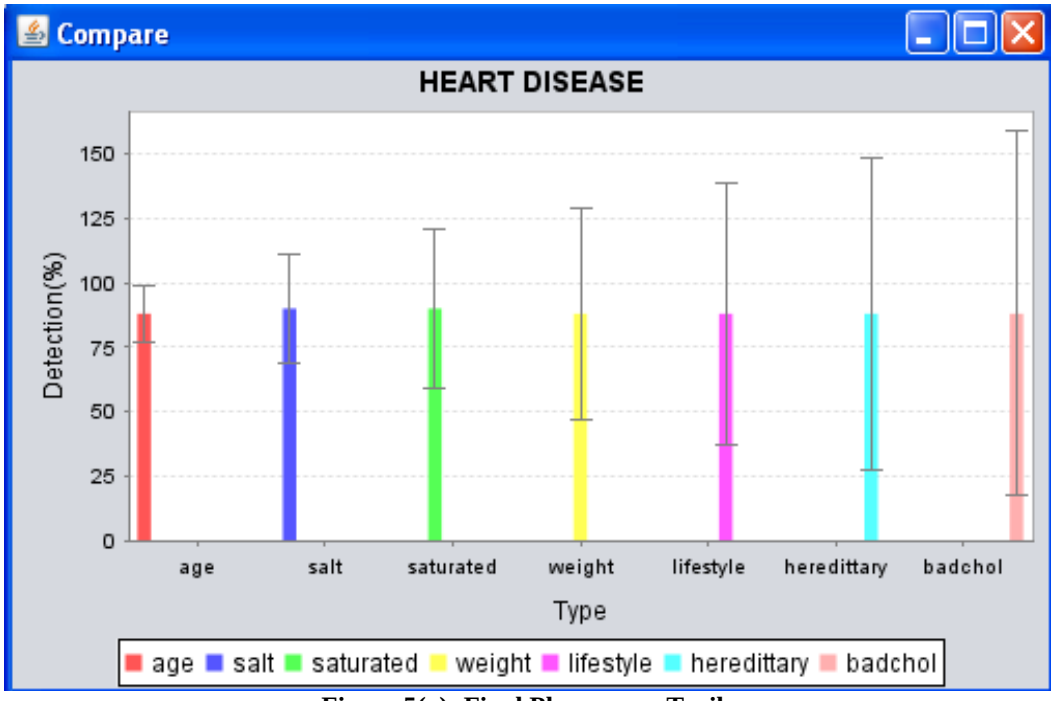

Figure 5(a): Final Pheromone Trails

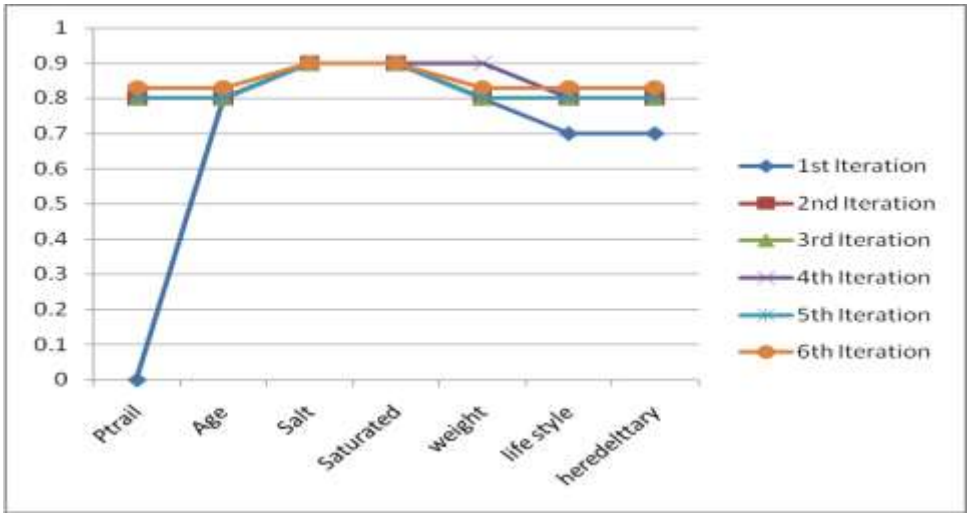

Figure 5. (b): Comparative analysis 1st to 6th iteration

Accuracy of Data

After applying different categorization as shown in the figure 4, we receive the continuous better accuracy in the case of increasing positive heart disease symptom. We provide the two best accuracy of our result which shows the betterment from the previous

Methodology adopted. The table 1 shows the accuracy. We also provide 3 different transactions and accuracy result by our approach.

Table 1: Accuracy of technique

Technique	Accuracy
ACO and Data Mining	91%
K-mean based MAFIA	85%
Naïve Bayes	80%

The features should be analyzed with the Java device. In this device, three different processes should be done with testing dataset. First of all, data should be tested with ACO technology by calling Processed Data on K-Mean Clustering. The results obtained are presented in the comparatively showing in figure 6.

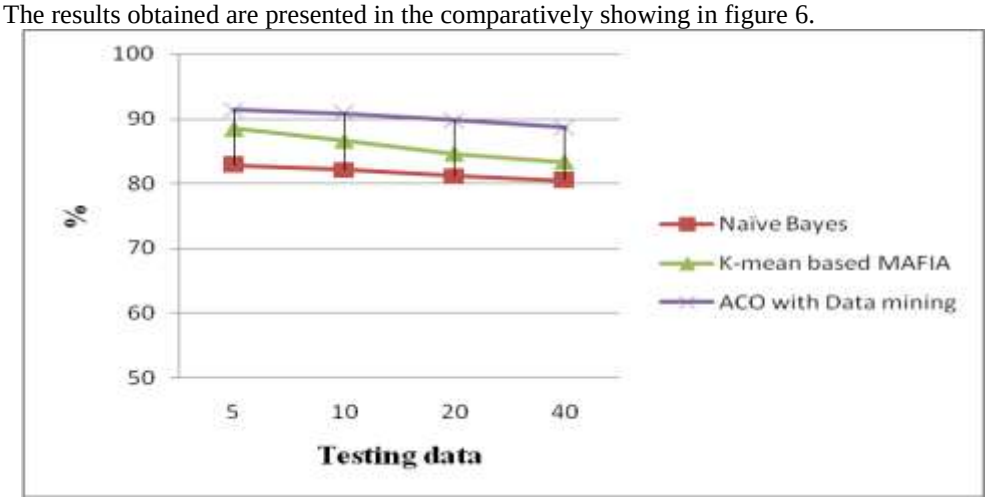

Figure 6: Comparison Accuracy for Naïve Bayes , K-mean based MAFIA and ACO with Data Mining

Heart Disease Attribute

We can also check the result by real-time evaluation. It is coded first in the data mining engine, and after debugging the program, we can check the output. In this way, we can check the diseases And risk in the patient according to the characteristics of the characteristics. If the value of attributes or inputs is high, then the patient has high risk and if the value or input is low, then the patient's risk of cardiovascular risk is low. And likewise, if the values are normal then the patient and the results show that the patient is normal. Here we are showing some examples that show high risk, low risk, etc. in the person

Table 2: Dataset Testing For Heart Disease Attribute

Chestpain	Cholestrol	Maximum heart rate	Blood Pressure	Blood Sugar	OldPeak	Result
0	100	70	60	50	0	0.07
0.1	155	78	74	83	0.1	0.1
0.16	158.5	83.2	78.9	123	0.2	0.25
0.5	250	110	130	150	0.5	0.5
0.1	281	121	150	161	0.7	0.62
0.89	353.5	136.8	190.9	231	0.866	0.77
1	400	150	185.3	235	0.92	0.8

We have tested the system designed with the following values for each region and data mining tool result. In the following table 2, we can see different values of different input variables and their results accordingly. If the values of inputs are in their lower categories then the risk is also low, resulting in a minimum value. Similarly, for higher values of input.

Performance Measures

The performance of data mining and ACO algorithms is measured using *specificity, sensitivity, and accuracy.* Due to its applicability in the field of medicine, it is considered as the most important. Illusion matrix is a type of table that provides a view to the performance of algorithms.

Table 3: Specificity, Sensitivity, and Accuracy Analysis

Method	Sensitivity (%)	Specificity (%)	Accuracy (%)
ACO with Data mining	90.71	91.29	91
K-mean based MAFIA	86.46	84.57	85.515
Naïve Bayes	81.15	80.39	80.77

Naive Bayes (NB) classifier: This is a straightforward classifier, which was probably based on the NB applied with the notion of strong independence.

K-mean based MAFIA: During this technique, K-mean filters the information from the database using the closest cluster

ACO - This is a customizable algorithm that is an algorithm to solve problems arising during ACO technology.

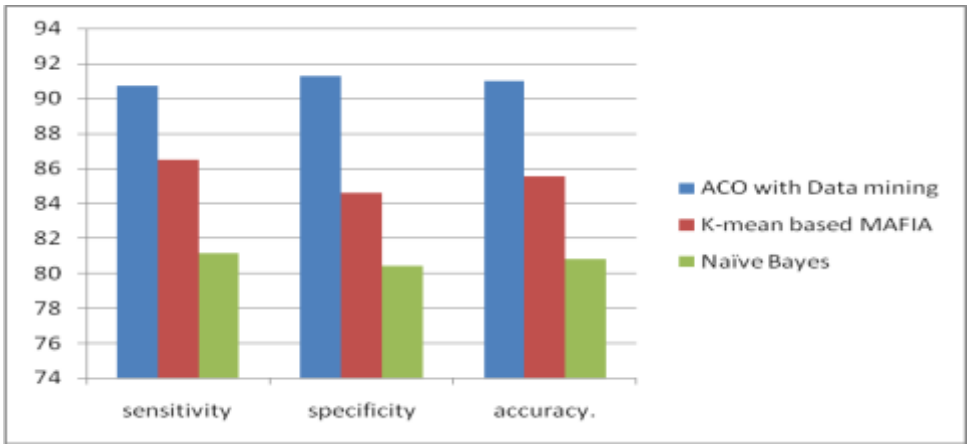

Figure 7: Specificity, Sensitivity, and Accuracy Analysis for different method

The data above for the results, training and test part in the above table 3 and figure 7 are obtained by dividing, which is randomly taken as stratified data points. Experimental results show that the value for the *specificity, sensitivity, and accuracy* of the ACO algorithm is high. Finally, the results show that ACO is most suitable for predicting heart attack compared to the K-mean based MAFIA and Naïve Bayes classification algorithms.

Conclusion

The use of data mining techniques in heart disease detection increases the likelihood of finding a right and initial diagnosis, which can prove to be important in fighting disease. We propose an efficient concept to detect cardiovascular diseases based on anti colony optimization and data mining. We use random ant generations for threshold settings and depending on that, we get global optimal value for each item set, so it's easy to find out. Then we apply the infection prospect for the next update of pheromone. Then it is rearranged when the pheromones are required. Important rules useful to predict the presence of heart disease were removed from the dataset. Domain expert confirmed that most of the rules generated are important in interpretation.

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. Wu R., Peters W., Morgan MW, "The Next Generation Clinical Decision Support: Linking Evidence to Best Practice", Journal Healthcare Information Management Vol 16(4), pp. 50-5, 2002.
4. Tsipouras, 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.
5. 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.
6. Fayyad U.M., G. Piatetsky-Shapiro, P. Smyth and R. Uthurusamy, "Advances in Knowledge Discovery and Data Mining", (AKDDM), AAAI/MIT Press, Massachusetts, February 1996.
7. Oded M. and Lior R., "Data Mining and Knowledge Discovery Handbook", Springer Science and Business Media Inc., New York,2005.
8. M. M. Gaber, A. Zaslavsky, and S. Krishnaswamy, "Mining data streams: A review," ACM SIGMOD Record, vol. Vol. 34, no. 1, June 2005.
9. J. Han, H. Cheng, D. Xin and X. Yan, "Frequent pattern mining: current status and future directions," Data Mining and Knowledge Discovery 2007, pp. 55-86.
10. 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.