Research Paper
Volume : 5 | Issue : 8 | Special Issue August-2016 • ISSN No 2277 - 8179

An Efficient Cardiovascular Disease prediction through Pheromone Based ACO with Hybrid Fuzzy logic

Engineering

KEYWORDS: Data mining, ACO, mining algorithm, cardiovascular diseases (CVDs), Fuzzy logic

S.Kavitha — Research Scholar, Vinayaka Mission University, Tamilnadu, India

Dr.M.Nithya — Head, Dept of CSE, VMKVEC, Vinayaka Mission University, Tamilnadu, India

ABSTRACT

World Heart Day is celebrated every year on 29th September with the intent of raising awareness about cardiovascular diseases. As per a World Health Organization (WHO) report, cardiovascular diseases (CVDs) would be the largest cause of death and disability in India by 2020. Changing lifestyle pattern, coupled with lack of physical exercise has put 74 per cent of urban Indians at risk of suffering from cardiovascular diseases (CVD), a recent study has revealed. Heart diseases have emerged as the number one killer for Indian women, they account for 15 % of the global burden of heart disease which kills about 15 million people every year. Currently, the key challenges that are faced by cardiac care in India are lack of awareness of non-communicable diseases by the people, inadequate facilities, accessibility etc. This research work deals with improving the accuracy of Heart Disease Prediction using Pheromone Based ACO with Hybrid Fuzzy logic. The overall objective of this research work is to design and implement ACO based data mining algorithm for Heart Disease Prediction. The proposed technique is designed and implemented in the MATLAB tool. The comparison among some well known technique has also been drawn by considering various performance metrics. We have evaluated our new classification approach using the well-known medical Cleveland, Abalone, Iris, Cardio to Cography, dataset as input. Results indicate that the proposed method can predict the Cardiovascular Disease with an acceptable accuracy. In addition, the extracted fuzzy rules have significant interpretability either.

I.INTRODUCTION

Data mining is the process of extracting the useful information from large amount of data. Data mining software is one of a number of analytical tools for analyzing data. Technically, data mining is the process of finding correlations or patterns among dozens of fields in large relational databases. The overall goal of the data mining process is to extract information from a data set and transform it into an understandable structure for further use. Data mining is the search for the relationships and global patterns that exist in large databases but are hidden among vast amounts of data, such as the relationship between patient data and their medical diagnosis. This relationship represents valuable knowledge about the database and the objects in the database, if the database is a faithful mirror of the real world registered by the database. ACO based Mining The Knowledge Discovery in Databases (KDD) field of data mining is concerned with the development of methods, techniques and algorithm which can make sense of the available data. Knowledge Discovery in Database is useful in finding trends, patterns and anomalies in the databases which is helpful to make accurate decisions for the future.

Recently, Ant Colony Optimization (ACO) algorithm has been applied to the data mining field to extract rule based classifier [8] [2].Ant Colony Optimization (ACO) is a metaheuristic for solving hard combinatorial optimization problems [6]. The inspiring source of ACO is the pheromone trail laying and following behavior of real ants, which use pheromones as a communication medium.

In analogy to the biological example, ACO is based on indirect communication within a colony of simple agents, called (artificial) ants, mediated by (artificial) pheromone trails [7] [9]. The pheromone trails in ACO serve as a distributed, numerical information, which the ants use to probabilistically construct solutions to the problem being solved and which the ants adapt during the algorithm's execution to reflect their search experience [4] [11]. In essence, the goal of data mining is to extract knowledge from data. Data mining is an inter-disciplinary field, whose core is at the intersection of machine learning, statistics and databases [7] [1]. We emphasize that in data mining – unlike for example in classical statistics – the goal is to discover knowledge that is not only accurate but also comprehensible for the user [1] [8]. Comprehensibility is important whenever discovered knowledge will be used for supporting a decision made by a human user [3] [1]. After all, if discovered knowledge is not comprehensible for the user, he/she will not be able to interpret and validate it. In this case, probably the user will not trust enough the discovered knowledge to use it for decision making. This can lead to wrong decisions [3] .

Cardiovascular Disease:

Facts:

It's no longer true that only those in their 50s and 60s can have cardiovascular problems. Even those in their 30s are coming with such problems these days.

Causes:

- lifestyle habits like junk food
- alcohol consumption
- smoking
- mental stress
- lack of physical work

Implication on Indian Health System:

(i) to provide information and an enabling environment for increasing awareness and adoption of health living habits by the community;

(ii) early detection of persons with risk factors and cost-effective interventions for reducing risk; and

(iii) early detection of persons with clinical disease and cost-effective secondary prevention measures to prevent complications.

II. LITERATURE REVIEW

Yumin Chen et al. [1] proposed a new rough set approach to feature selection based on Ant Colony Optimization (ACO), which can adopt mutual information based feature significance as heuristic information. The paper also proposed a feature selection algorithm. This research approach started from the feature core, which changed the complete graph to a smaller one. To verify the efficiency of this algorithm, experiments are carried out on some standard UCI datasets. The results demonstrated that this algorithm could provide efficient solution to find a minimal subset of the features.

Pablo Loyola et al. [2] proposed an ant colony optimization-based algorithm to predict web usage patterns. This methodology has incorporated multiple data sources, such as web content and structure, as well as web usage. The proposed model is based on a continuous learning strategy based on previous usage in which artificial ants try to fit their sessions with real usage through the

modification of a text preference vector. Subsequently, trained ants are released on to a new web graph and the new artificial sessions are compared with real sessions previously captured via web log processing. The main results of this work are related to an effective prediction of the aggregated patterns of real usage, reaching approximately 80%. This approach also allowed the obtaining of a quantitative representation of the keywords that in fluence then navigational sessions.

Md. MonirulKabir et al. [3] proposed a new hybrid ant colony optimization (ACO) algorithm for feature selection (FS), called ACOFS, using a neural network. A key aspect of this algorithm is the selection of a subset of the salient features of reduced size. ACOFS uses a hybrid search technique that combined the advantages of wrapper and filter approaches. In order to facilitate such a hybrid search, this research paper designed a new set of rules for pheromone update and heuristic information measurement. On the other hand, the ants are guided in correct directions while constructing graph (subset) paths using a bounded scheme in each and every step in the algorithm. The above combinations ultimately not only provide an effective balance between exploration and exploitation of ants in the search, but also intensify the global search capability of ACO for a high quality solution in FS. Extensive experiments were conducted to ascertain how AOCFS works on FS tasks. The comparison details show that ACOFS has a remarkable ability to generate reduced-size subsets of the salient features while yielding significant classification accuracy.

Kashef et al. [4] have proposed a novel feature selection algorithm based on Ant Colony Optimization (ACO), called Advanced Binary ACO (ABACO), is presented. The performance of proposed algorithm is compared to the performance of Binary Genetic Algorithm (BGA), Binary Particle Swarm Optimization(BPSO),Catfish BPSO, Improved Binary Gravitational Search Algorithm(IBGSA),and some prominent ACO- based algorithms on the task of feature selection on 12 well-known UCI datasets. Simulation results verify that the algorithm provides a suitable feature subset with good classification accuracy using a smaller feature set than competing feature selection methods. they present a new feature selection technique based on Ant Colony Optimization(ACO)by combining two models of ACO. The proposed algorithm has a strong search capability in the problem space and can effectively find the minimal feature subset.

Khalid M. Salama et al.[5] have explored the use of various classification quality measures for evaluating the BAN classifiers constructed by the ants. The aim of this investigation is to discover how the use of different evaluation measures affects the quality of the output classifier in terms of predictive accuracy. In their experiments, they use 6 different classification measures on 25 benchmark datasets. they found that the hypothesis that different measures produce different results is acceptable according to the Friedman's statistical test. They explored the effect of using 6 different classification quality measures for evaluating the candidate BN classifiers constructed by the ants and updating pheromone during the training phase of ABC-Miner.

Wen xiong et al. [6] have proposed a novel hybrid clustering approach, which uses adaptive ant colony optimization(ACO) to optimize the partition of data set, and utilizes enhanced particle swarm optimization (PSO) to refine the result of the adaptive ACO. Experiments displayed that the approach obtains smaller clustering evaluations on three data sets of University of California Irvine (UCI) and competitive results on two data sets of UCI, which verifying its availability.

Mr. K. Ravikumar et al. [7] have described a model that develops an ant colony algorithm for the discovery of spatial trend patterns found in a GIS traffic risk analysis database. The proposed ant colony based spatial data mining algorithm applies the emergent intelligent behavior of ant colonies to handle the huge search space encountered in the discovery of this knowledge. Genetic algorithm is deployed to evaluate the spatial risk pattern rule sets to its optimization on search phase in quick successions. The experimental results on a geographical traffic (trend layer) spatial database show that their method has higher efficiency in performance of the discovery process and in the quality of trend patterns discovered compared to other existing approaches using non-intelligent decision tree heuristics. The proposed results provide spatial decision trees for traffic risk patterns with optimized route structure with the ant agents.

Mohammad Saniee et al. [8] have presented an ACO to extract a set of rules for diagnosis of diabetes disease. The new presented algorithm uses ACO to extract fuzzy If-Then rules for diagnosis of diabetes disease, they call it FADD. They have evaluated their new classification system via Pima Indian Diabetes data set. Results show FADD can detect the diabetes disease with an acceptable accuracy and competitive or even better than the results achieved by previous works.

Yu-Min Chiang et al. [9]They utilizes some machine learning techniques and a meta-heuristic approach to classify cancers using microarray data. In the study, the ant colony optimization algorithm is introduced to select genes relevant to cancers, then the MLP and SVM classifiers are used for cancer classification. Experimental results show that selecting genes by using ACO can improve the accuracy of MLP and SVM classifiers. Besides, the optimal number of genes selected for cancer classification should be set according to the microarray dataset and gene selection methods. Although the proposed ACO gene selection algorithm improves the classification accuracy of MLP and SVM classifiers .

Shyi-Ching Lian et al.[10] Have proposed an extension of Ant-Miner is proposed to incorporate the concept of parallel processing and grouping. . Due to the algorithm design, Ant-Miner made a slight modification in this part which removes the parallel searching capability. Based on Ant-Miner, they propose an extension that modifies the algorithm design to incorporate parallel processing. For solving the classification rule problem, they design an algorithm with the concept of multi-level rule choosing mechanism in order to get more accuracy of rule induced. Furthermore, they provide a possible direction for researches toward the classification rule problem

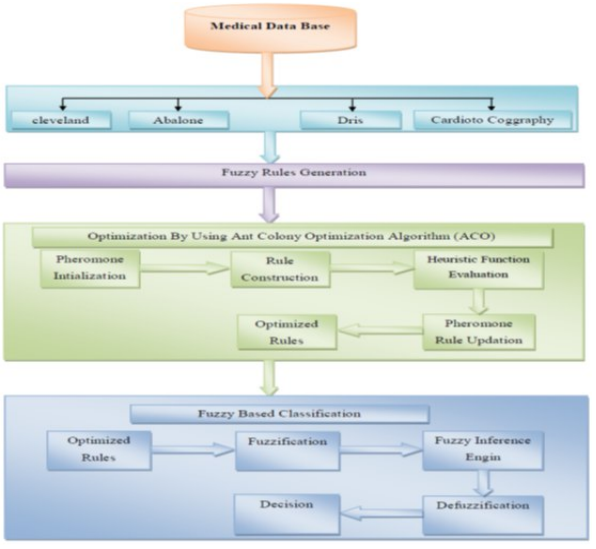
SurendraKumar, and C. S. P. Rao et al. [11] proposed a novel use of data mining algorithms for the extraction of knowledge from a large set of flow shop schedules. The purposes of this work is to apply data mining methodologies to explore the patterns in data generated by an ant colony algorithm performing a scheduling operation and to develop a rule set scheduler which approximates the ant colony algorithm's scheduler. Ant colony optimization (ACO) is a paradigm for designing Meta heuristic algorithms for combinatorial optimization problems. The ant algorithm is simple to implement and results of the case studies show its ability to provide speedy and accurate solutions. Further, this research has employed the genetic algorithm operators such as crossover and mutation to generate the new regions of the solution.

III. PROPOSED APPROACH

ACO can be used for clustering and classification in the field of data mining. In our proposed method, ACO will be used for pattern classification in which classification is based on fuzzy rules. The fuzzy rules are used to control the influence of pheromone values in ACO. We will describe a technique for update the pheromone rule that improves the quality of each rule. Because for each rule, the value of pheromone that increased in each iteration depend on the quality of modifications that ants will be done. With this new update pheromone function ants in order to improve the quality of rule, make better decisions in next iterations. The implementation is done in MATLAB with various mining datasets.The overall objective of this research work is to design and implement ACO based data mining algorithm for Cardiovascular Disease prediction

IV. RESEARCH METHODOLOGY

Initially, the input medical Cleveland, Abalone, Iris, Cardiotocography dataset is given to our proposed work. Based on these dataset we will generate the rules for fuzzy rules generation. By generating these fuzzy rules, the process repeatedly generates the same rules. To overcome this problem, we introduce the Ant Colony Optimization Algorithm (ACO), using this optimization algorithm we will generate the optimized rules. These rules are given in to the fuzzy that performs the fuzzification process then the rules are generated, then finally we perform the defuzzification process. Based on this process we will take the decision about the diseases. Finally it will give a accurate decision over the prediction model.



Block Diagram for Heart Disease Prediction

In order to predict the heart disease there are number of researches combined in fuzzy technique for effective classification. The main aim of this study is to predict the heart diseases because heart disease kills about 15 million people every year worldwide. Here numerous researches are combined in fuzzy technique with some other method but in our proposed fuzzy classification method an effectively classification process is performed.

4.1. Rules Generation

In the rules generation process fuzzy rules generation are used to generate the rules. Initially the input medical image dataset are given as input, the datasets are Cleveland, Abalone, Iris, Cardioto Cography.

4.2. Ant Colony Optimization Algorithm

The plan of an ACO algorithm implies the requirement of the subsequent features.

- A rule for pheromone updating, which precise how to fine-tune the pheromone trail (τ)
- A probabilistic transition rule based on the value of the heuristic function (η) and on the contents of the pheromone trail (τ) that is employed to iteratively erect a solution

4.3. Rule Construction

The rule is erected by the ant incrementally by including one term at a time. The term selection is based on the probability as specified by the equation 1.

$$p_{ij(t)} = \frac{\tau_{ij(t)} \cdot \eta_{ij}}{\sum_{i}^{a} \sum_{j}^{bi} \tau_{ij}(t) \cdot \eta_{ij}, \forall i \in I} \quad (1)$$

where η_{ij} is a problem-dependent heuristic value for $term_{ij}$, τ_{ij} is the amount of pheromone presently accessible (at time t) on the connection among attribute i and value I is the set of characteristics that are not so far applied by the ant in the domain of attribute i.

The rule pruning procedure iteratively eliminates the term whose elimination will cause a maximum increase in the quality of the rule. Using the subsequent equation (2) the quality of a rule is calculated.

$$Q = \left[\frac{TruePositive}{TruePos + FalseNeg}\right] * \left[\frac{TrueNegative}{FalsePos + FalseNeg}\right] \quad (2)$$

where True Pos is the number of cases wrapped by the rule and containing the similar class as that forecasted by the rule, False Pos is the number of cases wrapped by the rule and containing a dissimilar class from that forecasted by the rule, False Neg is the number of cases that are not wrapped by the rule, while containing the class predicted by the rule, True Neg is the number of cases that are not wrapped by the rule which have a dissimilar class from the class forecasted by the rule.

4.3. Fuzzy Logic

Fuzzy based classification is a method of generating a mapping from a given input to an output using fuzzy logic. Then, the mapping gives a basis, from which decisions can be generated. Membership Functions, Logical Operations, and If-Then Rules are used in the Fuzzy Logic based Process. The Stages of Fuzzy are,

1) Fuzzification
2) Fuzzy Inference Engine
3) Defuzzification

Fuzzification: Adapts the crisp input to a linguistic variable with the membership function accumulated in the fuzzy knowledge base.

Fuzzy inference engine: With the help of If-Then type fuzzy rules, alters the fuzzy input into the fuzzy output.

Defuzzification: Changes the fuzzy output of the inference engine to crisp using membership function equivalent to those exploited by the fuzzifier. Crisp rules are fuzzified inference system through the triangular membership function in our effort. Fuzzification is required as a degree of membership function is precised for each member of set. The fuzzy system predicts the results more specifically with the optimized membership function.

V. RESULTS AND DISCUSSION

The experimental result of fuzzy based classifier is discussed below. The proposed system is implemented using MATLAB 2014 and the experimentation is performed with i5 processor of 3GB RAM.

5.1 Dataset description

The proposed fuzzy based classifier is experimented with the three dataset namely Cleveland, Abalone, Iris, Cardio to Cography. These four datasets are taken from the UCI machine learning repository.

- Cleveland data[data set1]
- Abalone Data Set[data set2]
- Iris Data Set[data set3]
- Cardio to Cography[data set4]

5.2 Evaluation metrics

It consists of a set of measures that follow a common underlying evaluation methodology some of the metrics that we have choose for our evaluation purpose are True Positive, True Negative, False Positive and False Negative, Specificity, Sensitivity, Accuracy, F measure.

The test data outcome can be positive means predicting that the person has the heart disease or negative means predicting that the

person does not have the heart disease.

True Positive = unhealthy people correctly identified as unhealthy
True Negative = healthy people correctly identified as healthy
False Positive = unhealthy people incorrectly identified as unhealthy
False Negative = healthy people incorrectly identified as healthy

5.3 Performance Analysis

The performance of the proposed heart diseases prediction methods of human beings are evaluated by the six metrics Precision, Recall, F-Measure, Sensitivity, Specificity and Accuracy. The results of proposed work help to analyze the efficacy of the prediction process. The subsequent table II tabulates the results. Here, only the results of 4 datasets given in table II.

Table II: Results of the proposed heart disease prediction system with dataset

Sensitivity	Specificity	Accuracy	Recall	Precision	F_measure
0.7	0.8	0.9166	0.8888	1	3.5555
0.75	0.6	0.8695	0.8888	0.9411	3.5555
0.6666	0.6666	0.7352	0.7692	0.8695	3.0769
0.625	0.5	0.6888	0.7142	0.8620	2.8571

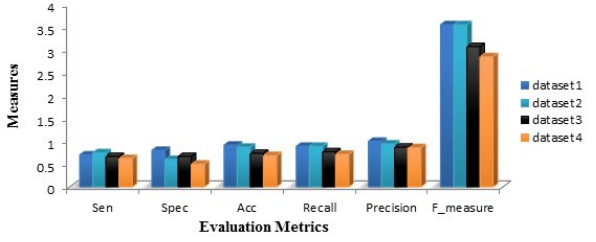


Fig. 5: Graph for dataset results with the Performance measures Precision, Recall, F-Measure, Specificity and Sensitivity, Accuracy

From table II, the evaluation metrics are analyzed for the four numbers of datasets, by which we can observe the efficiency of proposed detection system. The results of the measures Precision, Recall and F-Measure are graphically represented in fig. 5. The precision of four frames are 1%, 0.941176471%, 0.869565217%, and 0.862068966% and the average precision value for the whole frames of the dataset is 0.918203663%. The values of recall for the four datasets are 0.888888889%, 0.888888889%, 0.769230769%, and 0.714285714% and its average value is also 0.815323%. Likewise, the average F-Measure value for the entire dataset is 3.261294%. With these metrics, the specificity and accuracy are the main measures for evaluating the detection and tracking accuracy of our proposed system, which yields 0.641666% of specificity and 0.80260% of accuracy on average for the dataset. The results get high accuracy results on behalf of the reduced error rates in the proposed system. From the fig. 5 also, we find out the minimal value of error rates for the four dataset.

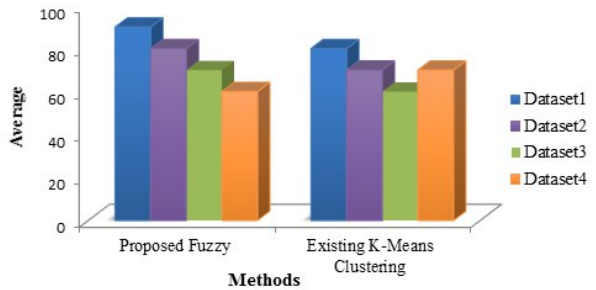


Fig. 12: Comparison of proposed method vs Existing method

VI.CONCLUSION

ACO can be applied to the data mining field to extract rule-based classifiers. The ant colony algorithm provides a suitable feature subset with good classification accuracy using smaller features.. The overall objective of this research work is to design and implement ACO based data mining algorithm for brain tumor detection. The proposed technique is designed and implemented in the MATLAB tool. The improved good accuracy outcomes of heart disease prediction are presented by our proposed work. In comparison with the K-Means Clustering gives very less accuracy values for the evaluation measures. The dataset 1 for the K-Means Clustering are 80% which is low in compared with our classifier, Fuzzy 90%. In dataset 2 for the K-Means Clustering are 70% which is low in compared with our classifier, Fuzzy 80%. In dataset 3 for the K-Means Clustering are 60% which is low in compared with our classifier, Fuzzy 70%. From these outcomes, it is known that by means of Fuzzy classifier in our work provides very good for the prediction purpose as it gives improved accuracy outcomes. T Thus, we can observe that our proposed work is better than other existing works for the heart disease prediction

REFERENCES

[1] Chen, Yumin, Duoqian Miao, and Ruizhi Wang. "A rough set approach to feature selection based on ant colony optimization." Pattern Recognition Letters 31, no. 3 (2010): 226-233.

[2] Pablo Loyola, Pablo E. Roman, Juan D. Velasquez, "Predictingwebuserbehavior-usinglearning-basedantcolony-optimization", Engineering Applications of Artificial Intelligence, Vol. 25, No. 5, pp. 889–897, 2012.

[3] Md. MonirulKabir, Md. Shahjahan, and Kazuyuki Murase, "A new hybrid ant colony optimization algorithm for feature selection", Expert Systems with Applications, Vol. 39, No. 3, pp. 3747–3763, 2012.

[4] Kashef, Shima, and Hossein Nezamabadi-pour. "An advanced ACO algorithm for feature subset selection." Neurocomputing 147 (2015): 271-279.

[5] Salama, Khalid M., and Alex Alves Freitas. "Investigating the impact of various classification quality measures in the predictive accuracy of ABC-Miner." In Evolutionary Computation (CEC), 2013 IEEE Congress on, pp. 2321-2328. IEEE, 2013.

[6] Xiong, Wen, and Cong Wang. "A novel hybrid clustering based on adaptive ACO and PSO." In Computer Science and Service System (CSSS), 2011 International Conference on, pp. 1960-1963. IEEE, 2011.

[7] Ravikumar, K., and A. Gnanabaskaran. "ACO based spatial data mining for traffic risk analysis." In Innovative Computing Technologies (ICICT), 2010 International Conference on, pp. 1-6. IEEE, 2010.

[8] Fathi Ganji, M., and Mohammad Saniee Abadeh. "Using fuzzy ant colony optimization for diagnosis of diabetes disease." In Electrical Engineering (ICEE), 2010 18th Iranian Conference on, pp. 501-505. IEEE, 2010.

[9] Chiang, Yu-Min, Huei-Min Chiang, and Shang-Yi Lin. "The application of ant colony optimization for gene selection in microarray-based cancer classification." In Machine Learning and Cybernetics, 2008 International Conference on, vol. 7, pp. 4001-4006. IEEE, 2008.

[10] Li, Pei-Chiang. "The Application of Ant Colony Optimization to the Classification Rule Problem." (2007)

[11] Surendra Kumar, and C. S. P. Rao, "Application of ant colony, genetic algorithm and data mining-based techniques for scheduling", RoboticsandComputer-Integrated Manufacturing, Vol. 25, No. 6, pp.901–908, 2009.