Application of Ant Colony Algorithm and Principal Components Analysis in the Diagnosis of Lung Cancer

Saeed Ayat[1], Mohsen Rahi[2,*]

[1] Associate Professor, Department of Computer Engineering and Information Technology, Payame Noor University, IRAN.
[2] M.Sc. student, Department of Computer Engineering and Information Technology, Payame Noor University, Iran
[1]Dr.ayat@pnu.ac.ir, [2] eng.rahi@gmail.com

Abstract

This paper presents a new method for diagnosing lung cancer by combination of ant colony algorithm, fuzzy logic and principal component analysis (PCA). In this method, PCA method is used to reduce the size of data sets, the fuzzy logic is used to create fuzzy rules that make it possible to be interpreted by experts. Finally, these fuzzy rules are optimized by ant colony algorithm (ACO). Evaluation and comparing the proposed method with other methods have been proposed to implement this approach, leading to lung cancer dataset with criteria such as speed, reliability, and the ability to interpret the show.

Keywords: lung cancer, data mining, ACO, fuzzy logic, PCA.

1. Introduction

Data mining usually means the tools and procedures for discovering new and suitable knowledge from databases and indeed it is a type of necessary knowledge discovery to problem solving in a determined domain. Due to the large volume of data and complexity of them, the need to effective and efficient tool in order to discover useful knowledge from this data is really required. Our goal is to discover and fetch the required knowledge using classification along with rule-based fuzzy systems which are a set of fuzzy if-then rules. Fuzzy rules are attractive because they can interpret by to the human expert [1, 2]. Diagnosis of various diseases in medicine is one of widely used ones in the fields of data mining and classification using fuzzy systems, and recently many studies have been done about it. In fact, the knowledge can be considered as a fuzzy rule base during the process of data mining and improve using optimization algorithms with regard to some criterions such as interpretability and accuracy. Interpretability doesn't mean when doctors are faced with a black box and simply present (though true) and offers no explanation. When different tests are done on a patient, the final diagnosis can be difficult even for a professional specialist. So, over the last

few decades, computer diagnostic tools used by physicians to remove irregularities from the data [3-5].

Lung cancer is the second most common cancer [6]. When lung cells start to grow uncontrollably and form a tumor, it is lung cancer. Lung tissues are involved in cancer in about 10 to 30 years before symptoms are known. Physicians are trying to determine whether the tested patient has a benign or malignant cancer. Because of the non-linearity classification in lung cancer, the development of a comprehensive model using conventional statistical-modeling techniques which included the all independent variables is very difficult. In addition, the combination of statistical techniques and traditional data management tools of the old data for analyzing a large collection of data is not enough. Because of the need to improve prediction accuracy in the diagnosis of lung cancer, researchers are trying more to apply artificial intelligence techniques to predict lung cancer [7-8].

Janecek and et al [9] examined the relationship between the classification and selection of properties. Moreover, they studied the effect of dimension reduction on classification by the principal component analysis (PCA). The results show that the classification accuracy based on the PCA is depending on the type of data and received changes on the components which are indicator of a basic requirement for the classification.

Duangsoithong and Windeatt [10] presented a method for reducing the data volume in the series which have high features and low samples as well as classification performance. By removing the irrelevant and repetitive data, new data sets are created. The results are more accurate and computation time is better. For example, for lung cancer 56 features reduced and 11 features achieved.

Azofra and et al [11] in a study about effective features of data sets, found a subset of available features to improve the performance of a learning algorithm. A large number of methods (based on searches) have been established for a set of features that can improve the function evaluation. Reliability is a well-known feature for a set of estimators. Based on the concept of reliability, a set of feature measurements is created for evaluating the feature set. Also how the proposed measure used to improve the search process (such as selecting of the most appropriate feature set) is shown.

In this paper the problem of diagnosing lung cancer is discussed. After a few preprocessing step, a combination of fuzzy systems and evolutionary algorithms is used to diagnose lung cancer, and a knowledge of input data sets is presented. The purpose of this study is to reduce the large size of the data sets with some preprocessing steps (PCA method) to select the optimum characteristics and then the operation of classification with the rule-based fuzzy systems which are set of fuzzy if-then rules, so the needed knowledge would be discovered and fetched. Applying the algorithm of the proposed method is shown in figure 1.
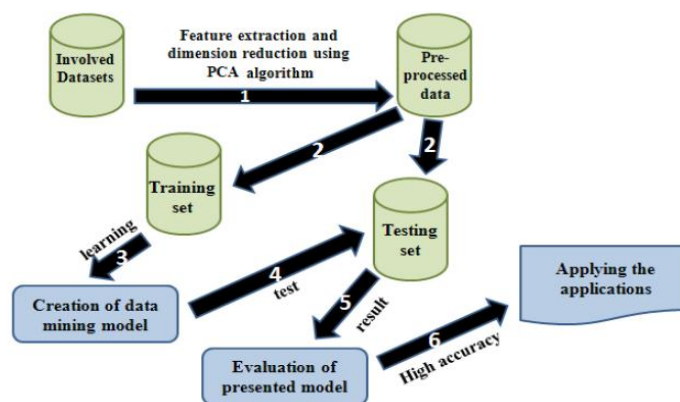

Fig 1: the processes of using algorithms of proposed method

Reducing the volume of the data sets is attractive, because the acquisition of the knowledge takes much less time with greater accuracy. The knowledge can be considered as a rule based [12] fuzzy that can be improved during the mining process with the optimization algorithms [13-14]. In order to optimize the set of fuzzy rules, the ant colony optimization algorithm is used. ACO algorithms can be used as a way to explore the rules in the field of data mining to make the best of the knowledge to be used [14, 15]. These algorithms (preprocessing and improving laws) are based on statistical concepts and evolution. The algorithm with the set of cooperation rules of fuzzy if-then rules in the phase state space to deal with, to have the best performance. Finally, the proposed system on a large volume of reservoir data mining University of California-Irvine cited by many researchers [16], have been applied.

We believe that the combination of fuzzy systems, PCA and ACO algorithms can be a powerful system for problems with a large search space that is more accurate and reliable compared to other methods.

2. Preprocessing using PCA algorithm

Principal components analysis (PCA) is one of the Data Reduction methods. This algorithm is a useful statistical technique that tries to analyze the major components of our data. The application of this conversion is in compression and reduction of dimensions as well as minimizing the error mean square of compression. If there is insufficient input data, PCA can extract the optimum conversion. If you have to choose the most important variable or a limited number of variables in a set, PCA can be used. PCA technique is the best way to reduce the size of the data linearly [17]. Removing the insignificant coefficients obtained from the conversion, data loss is lower than other methods. In this method, the new coordinate axes defined for data and the data are based on the new coordinate axes. The first axis should be placed in a way that variance is maximized (ie, in the direction that data dispersion is higher). The second axis must be perpendicular to the first axis in order to maximum the variance. Similarly the other dimensional axes perpendicular to all previous axes are placed in such a way that the data are for the highest dispersion. PCA is an optimal linear transformation as follows:

$$y = Wx \tag{1}$$

Where a conversion of data models with n dimension includes real values $(x \in R^n)$ on m dimension vector $(y \in R^m)$. Linear transformation matrix $W \in R^{m \times n}$ to obtain the maximum information is optimal and is calculated according to the following procedure:

$$\mu = E[x] = [E[x_1], E[x_2], ..., E[x_n]]^T \tag{2}$$

And the square covariance matrix is equal to:

$$R_{xx} = E[(x - \mu)(x - \mu)^T] \tag{3}$$

Where $E[.]$ is the operator of expected value. μ is mean vector of a data collection like x. Covariance matrix R_{xx} with real values defines the correlation between different x data features.

Covariance matrix R_{xx} that is $n \times n$, includes symmetric and true values $(R_{xx} \in R^{n \times n})$.

Eigenvalues of the covariance matrix R_{xx} have been arranged in descending order ($\lambda_1 = \lambda_{max}, \lambda_1 \geq \lambda_2 \geq ... \geq \lambda_n$). Imagine that the corresponding and orthogonal eigenvectors $e^1, e^2, ..., e^n$ form an orthogonal matrix $n \times n$ (these are vectors of length equal to one ($\|e\| = 1$).

$$E = \left[e^1, e^2, ..., e^n \right] \tag{4}$$

Choose m more important vector of the E matrix (the choice of particular vectors which have larger values) and transpose of them, the W matrix can be formed.

$$W = \begin{bmatrix} (e^1)^T \\ (e^2)^T \\ . \\ . \\ . \\ (e^m)^T \end{bmatrix} \tag{5}$$

PCA can be considered as an unsupervised learning from data can be. In fact, The PCA does not use of the available information about a class related to a pattern, but it discovers the correlation between the patterns, their elements and the main directions in which the patterns can be changed (with maximum Variance) see figure 2 [1].
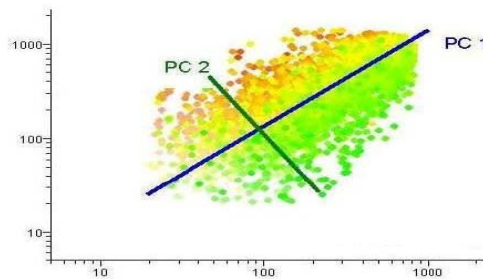

Fig 2: Principal components

3. Fuzzy rule systems

In this section, the method of knowledge representation using fuzzy rules and patterns classification are described. Then, the employed fuzzy model, making the initial fuzzy rules, determining the grade and degree of certainty for each rule are presented. Then, how to determine the class of a new sample and the used fuzzy reasoning are discussed.

3.1.   Patterns classification

Classification patterns are problems with n dimension, C grade and m educational model ($X_p = (x_{p1}, x_{p2}, ..., x_{pn})$, $p = 1,2,...,m$). Without the loss of the whole problem, each feature of the training modules is normalized in the number between [0, 1], (i.e., space of model for each feature is independently a number between 0 and 1. In this study, fuzzy if-then rules based on the established classification systems are used.

Rule R_j : *If* x_1 *is* A_{j1} *and* . . . *and* x_n *is* A_{jn} then   class C_j with CF_j , $j=1,2,..,N$ (6)

Rj : j[th] fuzzy if then rule

Aj1,... , Ajn: antecedent fuzzy sets on the unit interval [0, 1], like Large, Medium, Small and overlap each other.

Cj: achieved grade for each pattern.

 N: the total number of fuzzy if then rules.

CFj: the degree of certainty for Rj.

They used as initial fuzzy sets given from triangular fuzzy sets [13], [18].

3.2.   Fuzzy Model

According to Figure 3, the used fuzzy collection is a 5-part fuzzy collection. To encode it, each of these numeral areas is considered.
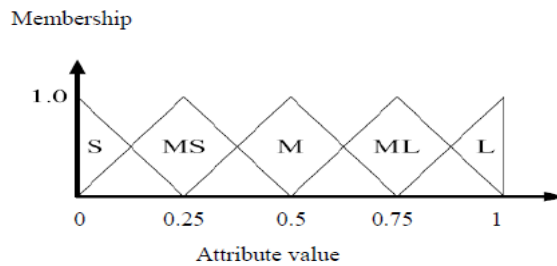

Fig 3: Proposed model for the partitioning of a fuzzy set

S (small) =1, MS (Medium Small) =2, M (Medium) =3, ML (Medium Large) =4,

L (Large) =5

3.3.   Initial fuzzy rules
Each rule is encoded by a series of values from 1 to 5. At first, a set of 100 rules, which each series is the length of the data series features, is generated randomly, and the diagnosis is analyzed.
In many similar functions, the initial rules are to be made directly from the training data. Although the time complexity of this method is better, but the method has just a high detection on training data, because the laws have not created with some good training. It is better to create initial rules randomly and always try to improve them [13], [15], [18].
How to calculate the respective class and the degree of certainty of the fuzzy if-then rules for optimize rules the fuzzy rule-based classification system, is as follows. For example, to determine the class (Cj) and the degree of certainty (CFj) of j rule, we have the following steps:
Step 1: Calculate the total adjustment degree $\beta_h(R_j)$ for h class (h= 1,….,C).

$$\beta_h(R_j) = \sum_{x_p \in Class\ h} \mu_{j1}(x_{p1}) \times ... \times \mu_{jn}(x_{pn}) \qquad\qquad h = 1,2,...,C \qquad (7)$$

Step 2: finding a class $(\hat{h})$ with the highest $\beta_h(R_j)$ value:

347

$$\beta_{\hat{h}}(R_j) = \max\{\beta_1(R_j), \beta_2(R_j),...,\beta_C(R_j)\} \qquad (8)$$

If more than one class have the highest value, R_j rule cannot be uniquely assigned to the C_j class. In this case the class of the rule is considered empty ($C_j = \phi$). If only one class is highest then $C_j = Class\ h$.

Step 3: If only one class has the maximum value for $\beta_h(R_j)$, then the degree of certainty for j rule (CFj) is determined with formula (9) :

$$CF_j = \frac{\beta_{\hat{h}}(R_j) - \overline{\beta}}{\beta_h(R_j)} \qquad , \overline{\beta} = \frac{\sum_{h \neq \hat{h}} \beta_h(R_j)}{C-1} \qquad (9)$$

3.4.   Fuzzy reasoning

With helping the way of producing a rule from previous section, n fuzzy rules can be randomly created. The rule belongs to the corresponding class and the degree of certainty of the all fuzzy if-then rules is determined. Class of a new sample (x) is defined by the following procedure ([13], [15], [18]):

Step 1: Calculation of $\alpha_h(x)$ for h class and h= 1,2, ..., C and the result is:

$$\alpha_h(x) = \max\{\mu_j(x) \times CF_j \mid C_j = Class\ h, j = 1,2,...,N\} \qquad , h = 1,2,...,C \qquad (10)$$

The

$$\mu_j(x) = \mu_{j1}(x_1) \times ... \times \mu_{jn}(x_n) \qquad (11)$$

Step 2: finding a class for the sample h_p^* with highest value $\alpha_h(x)$.

$$\alpha_{h_p^*}(x) = \max(\alpha_1(x),...,\alpha_C(x)) \qquad (12)$$

If more than one class has the highest value, then the sample x have not been classified. If not, the h_p^* class is assigned to the case x.

3.5.   Rules Assessment

After creating the set of rules, they must be carefully evaluated to determine their accuracy with final rules. In this step, the way for evaluation of each fuzzy if-then is examined. The competency of the rule $NCP(R_j)$ is considered as the number of samples which are classified properly by the following rule:

$$fitness(R_j) = NCP(R_j) \qquad (13)$$

The evaluation function is defined as follows:

$$EF(S) = W_{NNCP}.NNCP(S) + W_S.(|S|) + W_L.(Length(S)) \qquad , (W_{NNCP} + W_S + W_L = 1) \qquad (14)$$

In the above equation $NNCP(S)$ the set of rules which are classified improperly or are not classified by the rule, $NNCP(S)$ is calculated by the following equation:

$$NNCP(S) = m - \sum_{R_j \in S} NCP(R_j) \qquad (15)$$

$|S|$ is the size of a set or in other words, the number of fuzzy if-then rules in a set of rules. $Length(S)$ is The total length of the existing rules . W_{NNCP}, W_S and W_L are positive weight which are considered for these criteria. If the accuracy and ability to classification of the proposed model is very important for us, W_{NNCP} are considered much more than W_S and W_L. If simplicity, compactness and ability to interpret are very important for us, W_{NNCP} parameter is considered approximately equal to the sum of W_S and W_L. In implement of the weights, W_{NNCP}, W_S and W_L are considered 0.99, 0.005, and 0.005, respectively.

4.   Ant Colony Algorithm (ACO)

Ant colony algorithm (ACO) is a method that is applied to simulate the natural behavior of ants. The basic idea of this method was proposed by Dorigo [19, 20]. An ant colony system includes simple agents (ants) which are working together to generate a output as a single behavior of the entire system; in other words, the agents make a strong system which is able to find high quality answers for issues that have a large search space. The discovery rule, an ant colony system is able to do a strong and flexible search for a good combination of reasonable conditions which contain some predictable properties. Then, ACO as an optimal seeking idea to escape from local optimum answers and the convergence for global and optimal answer, were proposed [14]. In the following, natural ant colony system is analyzed and then the outline of the proposed algorithms is explained.

4.1.   Corresponding process of ACO and classification of fuzzy if-then rules

ACO algorithm parameters for classification of fuzzy rules:
•   Number of available ant in ant colony: number of rules used for classification.
•   Ant properties: attributes and parameters associated with the different range of data sets.
•   The amount of pheromone related to each ant: the amount of evaluation function of the ant.
•   Updating of pheromone related to the selected ant: optimization of classified fuzzy if-then rule set.
•   The current ant colony: the current rule set.
•   Disruption of current ant colony: a change in the current rule.
•   A new ant colony: a set of new rules.
•   Calculate the fitness of the new ant colony: Calculate the evaluation function for a set of new rules
•   Accepting new ant colony with a certain probability, if the acceptation is greater than current acceptation. Adopting a new code if the evaluation function value is higher than the current evaluation function of the rule set.
In the above process, first a colony of ants is generated randomly and the acceptation of the current population using the whole of acceptation and the remained pheromone are calculated fitness function value is equal to the number of samples which are classified correctly. Then one of these ants are selected randomly with a special probability. It changes one of its properties and updates the amount of pheromone and if the fitness function has been improved, this change will be informed to the other ants.

4.2.   Implement the Proposed Method

A schematic of the proposed algorithm has the following steps:
        1)   Applying PCA algorithm to reduce the dimensions of data, which includes
        the following steps:
                ✓   Step1: collecting some data and eliminate classes

✓   Step 2: set the data

✓   Step 3: calculation of the covariance matrix

✓   Step 4: calculation of the eigenvalues of the covariance matrix of eigenvalues

✓   Step 5: selection of the components of the feature vector

✓   Step 6: get the new data

2)   Normalization of the data

3)   Data fuzzy making

✓   Step 1: set the border of classification dynamically

✓   Step 2: set of classification with the degree of certainty

✓   Step 3: calculating the set of rules

✓   Step 4: calculate the degree of certainty of the rule

✓   Step 5: calculate the degree of compatibility from feature of a sample or a component of a rule

✓   Step 6: detection a class of a new sample with fuzzy production rules

- find the best rules to the new samples in each set
- find the class for new examples

4) Create a new set of the fuzzy if - then rules

5) Evaluate the cost of the current rule set using the evaluation function

6) Save the best set of rules which the best assessment value have been obtained from it and the storage of set of rules that are likely are the best state of the bad state and are used to escape from the local maximums.

7) Use of ACO algorithm to improve the rules

8) Calculate the evaluation value of the new rule set using the evaluation function

9) Repeat steps 6 to 8

10) Returns the best set of rules

Ant colony algorithm can do the best search for the best modifications during the run period. The degree of bias for ith change in front of the current rule (R_C) to A_j is as follows:

$$P_a(R_C,i,A_j) = \frac{\lfloor \tau(R_C,i,A_j) \rfloor \lfloor \eta(R_C,i,A_j) \rfloor}{\sum\limits_{u=1}^{5} [\tau(R_C,i,A_u)][\eta(R_C,i,A_u)]}$$
(16)

In Equation 16 τ is the pheromone value and η is an initiative possibility which is achieved by the following equations:

$$\eta(R_C,i,A_j) = \frac{N(i,A_j)}{\sum\limits_{v=1}^{5} N(i,A_j)}$$
(17)

$$N(i,A_j) = \sum\limits_{p=1}^{m} n_p(i,A_j)$$
(18)

$$\sum\limits_{p=1}^{m} n_p(i,A_j) = \begin{cases} 1 & \text{avg} \max\limits_{v=1}^{5} \{\mu_{A_v}(x_{pi})\} = \mu_{A_j}(x_{pi}) \\ 0 & otherwise \end{cases}$$
(19)

According to this detector, the highest probable linguistic value is calculated for the front section of fuzzy rule. After a few changes, ants updates a local pheromone according to the equation 20:

$$\tau(R_C,i,A_j) \leftarrow \tau(R_C,i,A_j) - \rho.(\tau(R_C,i,A_j) - \Delta\tau(R_C,i,A_j))$$
(20)

Where $0 < \rho < 1$ is evaporation parameter of the local pheromone and $\Delta \tau(R_C, i, A_j) = \tau_0$ where τ_0 is the initial pheromone value. When all the ants search completely, then the overall pheromone updating is done by equation 21:

$$\tau(R_C, i, A_j) \leftarrow (1 - \alpha) . \tau(R_C, i, A_j) + \alpha . \Delta \tau(R_C, i, A_j) \tag{21}$$

Equation 22 is equal to:

$$\Delta \tau(R_C, i, A_j) = \frac{fitness(R_N)}{N_{Classh}} \tag{22}$$

Where $0 < \alpha < 1$ is the evaporation parameter of overall pheromone and R_N implies on the new rule. The result of the algorithm would be the best change for the current fuzzy rule.

The proposed method is based on a data set of UCI University with 32 lung cancer samples, 56 features and 3 classes which in each class has tested 9, 13 and 10 samples respectively. The number of features is reduced using PCA method from 56 to 6 features. To implement the proposed approach, we use MATLAB software. To evaluate the proposed method, 10-CV techniques [21] have been used. According to this technique, the data set of lung cancer is divided into 10 equal parts which 9 parts are for training data set and one part for testing data set. Then the process is repeated 10 times in order to evaluate the all 10 parts of dataset. Table 1 presents the results of the proposed method implementation is evaluated by 10-CV technique.

Table 1: The results of implementing the proposed method with techniques 10-CV

Repeat	1	2	3	4	5	6	7	8	9	10
Result training	619.629	615.629	624.629	622.629	613.629	619.629	617.629	614.629	620.629	616.630
Result testing	68.7	68.7	69.7	67.7	69.7	67.7	69.7	68.7	67.7	66.69
Number of rules	20	20	20	20	20	20	20	20	20	20
average length	1.4	1.2	1.05	1.55	1.1	1.3	1.4	1.3	1.3	0.75

According to Table 1, the mean length is 1/235. Compared with C4.5 method [22] with an average length of 4.1 rules and the MLP [23] with an average length of 1/35, we concluded that the proposed method has high interpretation ability.

Table 2 shows that the accuracy on a test data set with K-NN method is the best accuracy in this proposed method. For the testing data, the accuracy of the proposed method is located after the method of SVM (support vector machine).

Table 2: The results of different methods

method	C4.5	K-NN	SVM	MLP	proposed method
Testing accuracy	94.56	96.99	96.56	95.42	96.99
Training accuracy	97.29	98.09	98.41	97.45	98.21

5. Conclusion

The proposed method was presented for the detection of lung cancer and it helps us obtain the knowledge of cancer samples from our input datasets. Moreover, according to the reliability

and the ability to interpret the rules, the shortest rules with higher interpretation and reliability ability, have identified. In the proposed method, we used PCA method to reduce the number of features and 10-CV technique for evaluation of the proposed method. Each feature was normalized between 0 and 1, separately. Using Algorithm ACO in the state of the problem, optimal rules were searched. Finally, we concluded that compared to other methods, the proposed method has the high reliability and the ability to interpret. Thus, according to these results, the proposed method can be used when there are so many features (dimensions) such as patients, image processing, and speech recognition.

References

[1]Krzysztof J. Cios, Witold Pedrycz, Roman W. Swiniarski & Lukasz A. Kurgan: "Data Mining A Knowledge Discovery Approach". Springer Science, New York, 2007.

[2]J. Kennedy, R.C. Eberhart, A new optimizer using particle swarm theory, in: Proc. of the Sixth Int. Symp. On Micro machine and Human Science (MHS'95), Nagoya, Japan (1995) 39–43.

[3]Abbass H.A. An evolutionary artificial neural networks approach for breast cancer diagnosis. Artif. Intell Med (2002).

[4] Sa Chou S.-M.; Lee T.-S; Shao Y.E; Chen I-Fei. Mining the breast cancer pattern using artificial neural networks and multivariate adaptive regression splines. Expert system with Applications (2004) 27, 133-142.

[5]Ohmann C.; Moustakis V.; Yang Q.; Lang K. Evaluation of automatic knowledge acquisition techniques in the diagnosis of abdominal pain. Artif Intell Med, (1996), 8, 23-36.

[6]As Ferlay J, Shin HR, Bray F, Forman D, Mathers C, Parkin DM. Estimates of worldwide burden of cancer in 2008: GLOBOCAN. Int J Cancer 2010; 127 (2010): 2893-917.

[7]Subhash C; Sikha Bagui; Kuhu Pal; Nikhil R.Pal. Breast cancer detection using rank nearest neighbor classification rules. Pattern Recog (2003) 25-34.

[8]Ta-Cheng Chen; Tung-Chou Hsu. "A GAs based approach for mining breast cancer pattern". Expert System with Applications 2005.

[9]Janecek, A.G.K., Gansterer, W.N., Demel, M.A., Ecker, G.F.: On the relationship between feature selection and classification accuracy. Journal of Machine Learning and Research. JMLR: Workshop and Conference Proceedings 4.(2008) 90–105.

[10]Rakkrit Duangsoithong and Terry Windeatt: Relevance and Redundancy Analysis Quinlan for Ensemble Classifiers. Springer-Verlag Berlin Heidelberg, 2009.

[11]Antonio Arauzo-Azofra, Jose Manuel Benitez and Juan Luis Castro: A feature set measure based on Relief. RASC, 2004.

[12]Carlos Andres Pena-Reyes; Moshe Sipper. A fuzzy-genetic approach to breast cancer diagnosis. Artif Intell Med, 1999.

[13]Ajith Abraham and ravi jain: Soft computing models for network intrusion detection system. In Classification And Clustering For Knowledge Discovery. Studies In Computational Intelligence. Springer, Netserlands. (2005) 190- 207.

[14]Rafael S.; Parpinelli; Heitor S.Lopes; Alex A.Freitas. An Ant Colony Algorithm for Classification Rule Discovery. Idea GroupPublishing , (2002) 190-208.

[15]Monmarche N.; Freitas,A.A. On data clustering with artifical ants. AAAI Workshop (1999) 23-26.

[16] http://archive.ics.uci.edu/ml/machine-learning-databases.

[17]H. Azami, M. Malekzadeh, S. Sanei,A New Neural Network Approach for Face Recognition based on Conjugate Gradient Algorithms and Principal Component Analysis, Journal of mathematics and computer Science 6 (2013) 166 – 175.

[18]Tomoharu nakashima and hisao ishibuchi: Using Boosting Techniques to Improve the Performance of Fuzzy Classification Systems in Classification and Clustering for Knowledge Discovery. Studies In Computational Intelligence. Springer, Netserlands.(2005) 146-157.

[19]M. Dorigo, and L.M. Gambardella. Ant Colony System: A Cooperative Learning approach to the traveling salesman Problem. Journal of IEEE Transactions Evolutionary Computation, Vol 1, No 1, (1996) pp: 53-66.

[20] M. Fiuzy, A. Qarehkhani, J. Haddadnia, J. vahidi, H. Varharam, Introduction of a Method to Diabetes Diagnosis According to Optimum Rules in Fuzzy Systems based on Combination of Data Mining Algorithm (D-T), Evolutionary Algorithms (ACO) and Artificial Neural Networks (NN),Journal of mathematics and computer Science 6 (2013) 272 – 285.

[21]Kohavi R.A study of cross study of cross-validation and bootstrap for accuracy estimation and model selection. In Proceedings of International Joint Conference on AI, (1995) 1137–1145.

[22]J.R., C4.5, Programs for Machine learning, Morgan-Kaufman, San Mateo, Ca 1993.

[23] S. Mitra and L. I. Kuncheva, Improving classification performance using fuzzy MLP and selective partitioning of the feature space. Fuzzy Sets and systems, vol, 70, no, 1, (2001).