JOURNAL OF PURE AND APPLIED MICROBIOLOGY, Aug. 2015. Vol. **9**, p. 183-191
SPECIAL ISSUE ON RECENT RESEARCH CHALLENGES IN BIO-MEDICAL APPLICATIONS

Big Data Analysis and Design of Intelligent System for Early Detection of Ovarian Cancer in Women

G. Ravi Kumar

Sri Ramakrishna Institute of Technology, Coimbatore, Tamilnadu, India.

(Received: 01 February 2015; accepted: 28 April 2015)

Big Data Analysis is the market trend today in research institutes and industries. The data analysis problems can be solved by the machine learning and data mining techniques. The data mining techniques in the health care system are used to discover the valuable knowledge, which help the physicians to treat the patients at the earliest. This paper uses the data mining technology such as feature selection and classification to develop a predictive model for ovarian cancer detection. A large data set is gathered and preprocessed. Rough set theory is used to discover the data dependencies and reduce the feature set contained in the dataset. The Hybrid PSO and ACO (PACO)is used to optimize the selected featuresto efficiently classify the ovarian cancer tumors, either malignant(Stage I, StageII, Stage III, Stage IV) or benign.The Classification task is performed by the Multi Layer Feed Forward Neural Network (MFFNN) and it is trained using the backpropagation algorithm with momentum. The performance of the system is measured in terms of classification accuracy, mean absolute error and root mean square error.

Key words: Ovarian cancer, Big Data Analysis, Feature Selection, Classification, Rough Set Theory, Particle Swarm Optimization (PSO), Ant Colony Optimization (ACO), Multi Layer Feed Forward Neural Network.

Cancer is one of the major disease in the world and in most cases it results in death. The ovarian cancer ranks fourth as a cause of women's deaths. The 70% of women with ovarian cancer are not aware of the disease until the disease has advanced, stage III or IV and it leads to high death rate (Di Wang, et, al., 2014)[15]. The most effective way to reduce ovarian cancer deaths is to detect it earlier. An accurate and reliable diagnosis procedure should be followed by physicians for early diagnosis.

Access to accurate, comprehensive, and timely relevant ovarian cancer data for detect cancer earlier and determine the effectiveness of treatment. Knowledge base system capture and store huge amounts of data elements in several format such as text, video, image (Niloofar Mohammadzadeh, et, al., 2014)[6]. Physicians find it challenge to make accurate decisions when it contains huge volume of data. A new generation of computational techniques and tools are developed to assist people in extracting useful information from the rapidly increasing volumes of different type of data. Intelligent Data Analysis is data analysis with artificial intelligence techniques, which can provide access to hidden information in large databases of clinical data in hospitals (P.Ramachandran, et, al., 2014)[8], (K.K. Ramesh Kumar, et, al., 2014)[9] . Intelligent data analysis with potential information definetly have a important role in prevention, treatment and accuracy in diagnosis (Zakaria Suliman, et, al., 2014)[16].

This paper utilizes the data mining technique to obtain a knowledge based system for ovarian cancer detection from an unorganized data set. The major contribution of the paper is as

* To whom all correspondence should be addressed.
E-mail: raviresearch14@gmail.com

follows: (i) initially, the rough set is applied for data reduction and to improve feature selection(ii) The Particle Ant Swarm optimization is used to optimize the rough set feature reduction to effectively classify the ovarian cancer tumors either malignant or benign. iii) The multilayer feed forward neural network is used for classification and it is trained with the optimized feature subset, where the different stages of ovarian cancer and non ovarian cancer data are classified.

The rest of the paper is organized as follows: Section II describes the recent related works about the system. Section III describes Preliminaries for this paper. The materials and methods are described in the Section IV. Experimental results and analysis of the proposed system is described in Section V. Finally, Section VI renders the conclusions.

Related Work

(Hesham Arafat, et, al., 2012)[3] investigates the strategies based on rough set theory with Particle Swarm Optimization for an adaptive feature selection to improve the quality of clinical diagnosis. The relevance selected feature subset by the hybrid approach is used to generate decision rules for the breast cancer classification task to classify the benign from the malignant cases. The classification accuracy of the proposed system is improved by the determined robust feature subset.

(Syed Imran Ali and Waseem Shahzad 2012)[10] proposed a feature selection algorithm based on conditional mutual information and ant colony optimization inorder to improve the classification accuracy. The algorithm is pure feature subset selection, which attain less computational cost and improved classification accuracy. The experiment is done on the thirteen benchmark datasets over the wellknown classification algorithms. The results shows the effectiveness of the proposed algorithm.

(Selva Mary, et., 2014)[11] proposed approach with fuzzy inference system for effective classification of genes to their matching gene types. The proposed classification technique uses the effective classification techniques such as Naïve Bayes classifier, k-NN and SVM. The popoular dimensionality reduction such as MPCA, PCA based dimensionality reduction, where the important genes are identified and extracted and an improved classifier is used for classification. The experimental results of the system perform better satisfactory results, while classifying the micro array gene expression dataset.

(Dr.K. Usha Rani, 2010)[12] proposed a parallel approach by using the feed forward neural network to help the diagnosis of breast cancer. The breast cancer data set is trained with back propagation algorithm combined with momentum and variable learning rate in the neural network. The experiment is compared with the multi layer neural network and single layer neural network, and the results shows the multilayer network is trained quickly than single layer neural network and also provides satisfactory results for the classification task.

(Di Wang, et, al., 2014)[15], proposed a novel self organizing neural fuzzy inference system to help the diagnosis of Ovarian cancer. Genetic Algorithm based rough set clustering incorporated Neural fuzzy inference system, which make use of the inference rule base automatically derived by Genetic algorithm based rough set clustering technique. The different data set is collected for the experiments and the results shows the system attain better accuracy and more reliable

(Tuan Zea Tan, et, al., 2008)[13], proposed a method for early detection of ovarian cancer to reduce the death rate. The Complementary learning fuzzy inference neural network(CLFNN) is used to exploit the lateral inhibition between the negative and the positive samples. The CLFNN is tested against the blood test data set, micro array data set and proteomics data set and it improves the diagnosis accuracy with higher consistency.

(Bichen Zheng, et, al., 2013)[2], proposed a methodology for feature extraction to help for breast cancer diagnosis. A K-means and SVM is combined and an algorithm is developed. The hidden patterns of the benign and malignant tumors are recognied separately by utilizing k-means algorithm. After that the SVM is used to obtain a classifier for differentiating the incoming tumors. The method is tested against the Wisconsin Diagnostic Breast Cancer (WDBC) data set and the accuracy of the result attains 97.38%.

Preliminaries

Rough Set Theory

In early 1980, the rough set theory was introduced by pawlak, which is a mathematical tool to deal with uncertainty. Using the data alone, the

roughest theory determines the data dependencies and reduce the number of attributes(P. K. Nizar Banu and H. Hannah Inbarani, 2012)[5] (Amit Saxena, et.al., 2014)[1]. The rough set helps to reduct the irrelevant attributes with minimal information loss.

Let Information system IS =(U,A), where U is a non-empty finite set of condition attributes, $\forall a \in A$, and there exists a corresponding function, $f_a: U \rightarrow V_a$, where V_a is the set of values of a. A reduct of attributes set of condition attributes A is a minimal set of attributes $B \subseteq A$, where all attributes $a \in A - B$ are dispensable and an associated equivalence of indispensability relation indicated by IND (B)

$$IND(B) = \{(x,x') \in U^2 | \forall a \in B a(x) = a(x')\} \quad ...(1)$$

Reducts are the subsets that are minimal, which does not contain any dispensable attributes. The set of all reduct of an IS is denoted by RED(IS). In feature selection process, the reduct calculation has great significance and it enables the evaluation of absolute reduction as well as reduction with core (P. K. Nizar Banu, *et al*., 2012)[5].

MATERIALS AND METHODS

The goal of this paper is to develop a knowledge base system to efficiently classify the different stages of ovarian cancer and normal case. The dataset contains a numerous number of attributes, to discover a minimum set of relevant attributes to describe the dataset, a feature subset selection should be performed. The rough set theory is well known technique for feature subset selection. The Hybrid Particle Swarm Optimization and Ant Colony Optimization (PASO) is used to optimize the rough set feature reduction to effectively classify the ovarian cancer tumors either malignant (Stage I, Stage II, Stage III, Stage IV) or benign. The Multi layer Feed forward neural network (MFFNN) is used in the classification stage to classify the data into different ovarian cancer stages and normal case. The figure 1 shows the proposed intelligent system for early detection of ovarian cancer.

Data Source

The Data set has been taken from the Singapore National University Hospital (NUH)[17] where the dataset based on the blood test results of 172patients, where the set contains 23 of normal and 78 of benign cyst, 10 of borderline, 19 of Stages I and II patients and 42 of Stages III and IV patients. The data set used for this research consists of 28 features and the feature are listed in the table 1.

PASO and Rough set- based feature Selection

The PASO is used for the optimal feature selection process. A large feature space is considered, which consists of feature subsets and each feature subset is viewed as a position in such a space. There are 2N kind of subset for N feature and they are different from each other in the length and features contained in each subset. The subset with highest classification quality and least length are considered as optimal position. The particle

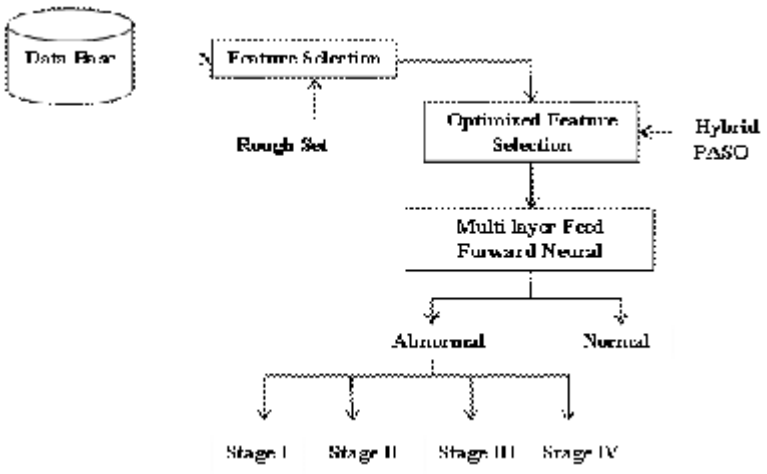

Fig. 1. Intelligent system for early detection of ovarian cancer

swarms are placed in the feature space, where each particle occupies one position. The i^{th} particle is represented as $P_i=(p_1,p_2,.....p_n)$. The best position of any particle is $X_i= (x_1,x_2,.....x_n)$. The velocity of particle is represented as $V_i= (v_{i1},v_{i2},.....v_{in})$. The Particle flies to the space to determine the best position. The particle changes their position by communicating with each other for a time and search around the local best and global best position. The ants are placed in the space equal to the number of particles to efficiently perform global exploration for quickly attaining the feasible solution. A solution willbe generated by each ant around the global best position found among all particles in the swarm up to iteration count. The pheromone guided search mechanism of an ant colony is implemented locally to synchronize the position of the particle in the PSO to attain the feasible domain of the objective function faster. The exploration ability of particle swarms guided with ant has a better ability to perform feature selection and discovers optimal subsets(Parag Deoskari , et, al., 2013) [7].The Figure 2 shows a feature selection algorithm using PASO and Rough Set

Table 1. Imporatant feature to identify ovarian cancer

Attribute Number	Attribute
1	Age
2	Blood Platelet
3	Packed Cell Volume (PCV)
4	Hemoglobin (Hgb)
5	²- Thromboglobulin (BTG)
6	Reaction Time of Thrombelastography (TEG-R)
7	Coagulation Time of Thrombelastography (TEG-K)
8	Maximum Amplitude of Thrombelastography (TEG-MA)
9	Fibrinogen
10	Factor VII
11	Von Willebrand-Factor(VMF)
12	Thrombin-Antithrombin Complex (T/AT)
13	Prothrombin Fragment 1&2 (F1+2)
14	Antithrombin III Activity (ATAC)
15	Antithrombin III Antigen (ATAG)
16	Plasminogen
17	D-dimer
18	Tissue Plasminogen Activator (tPA) Activity
19	Tissue Plasminogen Activator (tPA) Antigen
20	Urokinase-like Tissue Plasminogen Activator (uPA) Activity
21	Urokinase-like Tissue Plasminogen Activator (uPA) Antigen
22	Urokinase-like Tissue Plasminogen Activator (uPA) Receptor
23	Plasminogen Activator Inhibitor (PAI) Activity
24	Plasminogen Activator Inhibitor (PAI) Antigen
25	Protein C Antigen (PCAg)
26	Tissue Polypeptide Specific (TPS) Antigen
27	Tissue Factor Pathway Inhibitor (TFPI)
28	Cancer Antigen-125 (CA-125)

Representation and Updation of Velocity and Position

The PASO is applied by considering the position representation, velocity representation and position update strategies. The ***position*** is represented by binary bit strings of length N (total number of attributes. Every bit represents an attribute, the value '1' indicates that the corresponding attribute is selected, while '0' not selected. The each position represents an attribute subset. A positive integer is used to represent the ***velocity*** of each particle, which varies between 1 and V_{max}. It indicates how many of the particle's bits should be changed, at a particular instance in time, to be similar as that of the global best position, i.e. the velocity of the particle flying on the way to the best position. The number of different bits between two particles describes the difference between their positions. Each particle velocity is updated as following equation.

$$V_{ik} = w \times v_{ik} - c_1 \times rand(\quad) \times (x_{ik} \quad p_{ik}) \quad c_2 \times Rand() \times (x_{gk} \quad p_{ik}) \qquad ...(2)$$

where c_1 and c_2 are acceleration constants, w is the interia weight and rand is to change the bits randomly.

The interia weight reduce the search area of the particle dynamically using

$$w_t = (w_{max} - w_{min}) \times \frac{maxiteration - t}{maxiteration} - w_{min} \qquad ...(3)$$

The particle update their position according to the equation (1) and (2). Consider xg be the number of different bits between the current particle and gbest and the new velocity be V. For updating the positions there exists two cases: i) $V \leq xg$ then the V bits of the particle are changed

randomly, different from that of the gbest. The particles move towards the global best instead of being same as gbest. Ii) $V > xg$then change all the different bits to be same as that of gbest and further randomly change $(V - gx)$ bits outside the different bits between the particle and gbest. After reaching the global best solution by the particle, it will keeps on moving towards other directions, enabling further search.

Fitness Function

Fitness function represents the objective function value of particle p at position x and calculated as

$$F(x_t^i) = \alpha \times \gamma_{x_t^i}(D) + \beta \times \frac{|C|-|x_t^i|}{|x_t^i|} \quad ...(4)$$

Input: C, the set of all conditional attributes
Output: Reduct R
Step 1: Intialize maximum iteration, random position and random velocity, c_1, c_2.
Step 2: For i=1 to S // S is the number of particles
Evaluate the interia weight for particle I by using equation (3)
Evaluate objective function for particle I by using equation (4)
End for
Step 3: Determine the $f_t^{best}(p_t^{best}) = \max\{f(x_t^1), f(x_t^2), \ldots . f(x_t^{i'})\}$
Step 4: Intialize $p_t^g = p_t^{best}$ and $Gbest_t - f_t^{best}(p_t^{best})$
Step 5: While (t< maximum iteration)
For i=1 to S
Calculate velocity v_t^i for particle I by equation (2)
Update particle position
Calculate the objective function for particle I by using equation (4)
Calculate pheromone based search solution a_t^i for particle i
Evaluate the objective function for $f(z_t^i)$
if $f(a_t^i) < f(x_t^i)$
$f(x_t^i) = f(a_t^i)$
$x_t^i = z_t^i$
End if
if $pbest_t^i < f(x_t^i)$
$pbest_t^i = f(x_t^i)$
$p_t^i = x_t^i$
End if
Determine the $f_t^{best}(p_t^{best}) - \max\{p_t^1, p_t^2, \ldots . p_t^p\}$
if $pbest_t^i < f_t^{best}$
$p_t^g = p_t^{best}$ and $Gbest_t = f_t^{best}(p_t^{best})$
End if
End For
End While
Step 5: Output reduct R

Fig. 2. Feature Selection algorithm

Where x_t^i is the condition attribute set and it is relative to decision D, $\gamma_{x_t^i}(D)$ is the classification quality, || is the '1' number of the length of the selected feature subset or the number of attributes for particle , |C| is the total number of features. The classification quality and the subset length depend on the two parameters and .

$\alpha \varepsilon [0,1]$ ...(5)

$\beta = 1 - \alpha$ ...(6)

The equation (5) and (6)indicate that the classification quality and feature subset length have various significances for the feature selection task. Higher á assures that the best position is at least a real rough set reduct. The fitness function is evaluated for each position and the criteria are to maximize the fitness.

Phermone guided search by ACO in PSO

The ant will be put into the solution space and it is equal to the number of particles in solution space. The pheromone guided search is applied by the ant after the initial optimization and it is used to update the position determined by the particle in the previous stage. Around the gbest position (determined among all particles in the swarm upto iteration count t, each ant i generate a solution (as the following equation

$$a_t^i = N(p_t^{ij}, \sigma) \quad ...(7)$$

where is the mean and is the standard deviation

Table 2. Reduced Feature Subset

Attribute no	Attribute
12	Thrombin-Antithrombin Complex (T/AT)
13	Prothrombin Fragment 1&2 (F1+2)
17	D-dimer
18	Cancer Antigen-125 (CA-125)

Table 3. Number of reducts using Different feature Selcetion Methods

Feature Selection method	Number of Reducts
PASO and Rough Set	4
PSO and Rough Set	7
Gentic And Rough Set	8
Rough Set	12

In equation (7), the algorithm generates component of the solution vector, which satisfy Guassian distributions with and , where initially at t=1 value of and at end of each iteration it will be updated as , where d is a parameter to produce a better position within this boundary. The objective function value of is computed by the equation (1) by replacing the position of (Current position of particle i). if F() < F then = and . The pheromone guided search technique of ant colony is integrated with PSO to act locally to synchronize the position of the particles in solution space to obtain the feasible domain of the objective function (P.S. Shelokar, et., 2007)[12].

Classification

The feature space dimension is reduced and the data set with new optimized features has been built and it has been shown in the Table 2, which is related to the ovarian cancer. The MFFNN is used for the Classification to classify the data among different stages of ovarian cancer and non ovarian cancer cases. The Multilayer Feed forward neural network uses a supervised learning algorithm to predict what category the pattern belongs.

The input features will be fed into each input unit in the input layer and the value coming out of an input unit is labelled as x_j. A weight will be associated with output of each input node. The weight is represented as w_{hj}, where the weight of the input x_j goes to the hidden unit z_h. The weighted sum of inputs of each hidden node will be calculated by using the following equation

$$\sum_{j=0}^{d} w_{hj} x_j \quad ...(7)$$

Where d represents the number of input units.

The sigmoid function is applied after finding the weighted sum to determine the output of the hidden node and it is given in the following equation.

Table 4. Error values for an original feature set and reduced feature set

Classifier	Full Feature Set		Reduced feature Set	
	MAE	RMSE	MAE	RMSE
MFFNN	0.28	0.47	0.19	0.38
Naïve Bayes	0.31	0.55	0.25	0.48
SVM	0.29	0.49	0.25	0.44

$$z_h = sigmoid\left(\frac{\sum_{j=0}^{d} w_{hj} x_j}{1+e^{-\sum_{j=0}^{d} w_h x_j}}\right) \quad ...(8)$$

Where the w_h is calculated using the equation by changing the higher limit d to H, where h is the number of hidden nodes.

In MFFNN for classifing more than two classes, it will have K output nodes for K classes,then the network should learn more weights and it tends to have more expressive space to find a sigmoid function. At each output nodes, the sigmoid function is applied. The output node which generates the largest value tells that which class the given input belongs and it can be done by applying the max function to the outputs. The softmax function is applied to the weighted sum, where the softmax function has the ability to make the maximum value of the output to be close to 1 and rest of the output nodes to be close to 0 (Lynne E. Parker, 2006)[4]. The softmax function for the output node y_i is calculated by the following equation.

$$y_i = Softmax(o_i) = \frac{e^{oi}}{\sum_{i=1}^{k} e^{oi}} = \frac{\sum_{h=0}^{H} v_{ih} z_h}{\sum_{i=1}^{k} e \sum_{h=0}^{H} v_{ih} z_h} \quad ...(9)$$

Where v_{ih} is the weightgoing in to the output unit i from the hidden unit h.

In learning phase, the feed forward backpropagation algorithm is used. The gradient descent is used in the backpropogation algorithm for updating the weights to minimize the squared error between the network output and the target output values. The update rules are obtained by applying the partial derivative for the error function with respect to the weights to find how much each weight is contribute to the error. The process will be done for each layer iteratively in the network by starting with the last set of weight and working back towards the input layer.The weight update for each unit is given in the following equation

$$\Delta v_{ih} = \eta(r_i^t - y_i^t) z_h^t \quad ...(10)$$

$$\Delta w_{hj} = \eta\left(\sum_{i=1}^{k}(r_i^t - y_i^t)\right) z_h^t (1 - z_h^t) x_h^t \quad ...(11)$$

Where $r^t$is the target output, x^t is the input, y^t is the actual output and η is the learning rate and it usually set to 0.1.

The gradient descent takes a long time to converge. So the Momentum is used in the learning process, which update the current weight by considering the previous weight update (Dr.K. Usha Rani, et, al., 2010) [14]. By adding momentum, the new weight update equations become

$$v_{ih}^t = v_{ih}^t + \Delta v_{ih}^t + \alpha \Delta v_{ih}^{t-1} \quad ...(12)$$

$$w_{hj}^t = w_{hj}^t + \Delta w_{hj}^t + \alpha \Delta w_{hj}^{t-1} \quad ...(13)$$

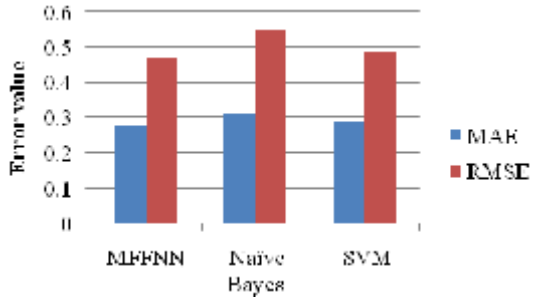

Fig. 3. shows the Mean absolute error and root mean square error for a) original feature b) reduced feature

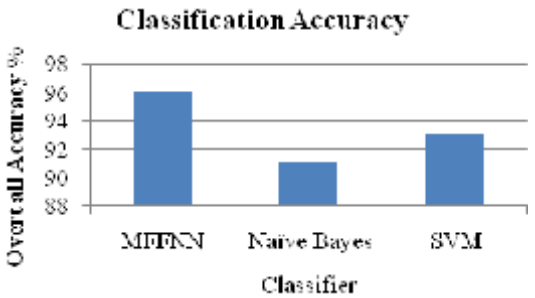

Fig. 3. Classification Accuracy

Table 5. Accuracy for different Ovarian Cancer Stages and Normal

Classifier	Normal	Stage I	Stage II	Stage III	Stage IV	Overall
MFFNN	100	98	95	95	92	96
SVM	94	92	90	88	85	91
Naïve Bayes	97	94	92	91	89	93

EXPERIMENTAL RESULTS

The dataset contains the 28 features, among them 4 features are selected by the PASO and rough set based feature selection, where these features have the ability to determine the seriousness of the tumor in ovary. These features are fed into the MFFNN and network is learned by backpropagation algorithm with momentum using these features. The Table III shows the number of features selected by the different methods and the PASO and rough set based feature selection has better ability to obtain an optimized feature set for classifying the ovarian cancer disease from the normal case.

The Performance of the feature selection algorithm is evaluated using the Mean Absolute Error (MAE) and the root mean square Error (RMSE).

The Mean absolute error is the average of the difference between the predicted and actual value in all test cases.

$$MAE = \frac{|a_1 - c_1| + |a_2 - c_2| + \cdots + |a_n - c_n|}{n}$$

...(14)

The root mean squared error yields the error value the same dimensionality as the actual and predicted values.

$$RMSE = \sqrt{\frac{(a_1 - c_1)^2 + (a_2 - c_2)^2 + \cdots + (a_n - c_n)^2}{n}}$$

...(15)

Where the c_i is the computed value and a_i is the corresponding corrected value

Table IV gives the MAE and RMSE value for original and reduced feature sets. These values are shown graphically in Figure 3. The MFFNN yields minimum error rate than other classifier with the reduced features. The performance of the PASO and rough set based feature selection attains minimum error rate for all classification techniques than the original feature set.

The accuracy of the proposed intelligent system is measured in terms of classification accuracy. The classification accuracy depends on the correctly classified instances and it is calculated by the following formula

$$Classification\ Accuracy = \frac{Number\ of\ instances\ classified\ correctly}{Total\ number\ of\ instances}$$

...(14)

The classification accuracy of the proposed system is compared with the SVM and Naive Bayes and it has achieved 96% accuracy overall and it is shown in the figure 3. The Table V shows the accuracy of the different stages of ovarian cancer and normal classified and the proposed system achieves better accuracy than the SVM and Naïve Bayes.

CONCLUSION

Ovarian cancer diagnosis is an important research because early detection and accurate staging will help to increase the survival rate of the patient. The paper proposed an intelligent system based on the data mining Feature Selection and Classification. The Feature Selection using the Rough set theory and obtaining an optimized feature set by combining PSO and ACO for better convergence. The Multilayer Feed Forward Neural Network is used for classifying the different stages of ovarian cancer and normal cases by using the optimized features. The Backpropagation algorithm with momentum is used on the learning stage for the classifier. The experimental results show that the proposed intelligent system attain better accuracy than other classifier technique using the reduced features. The proposed Intelligent system is capable to provide a suggestion for the diagnostic processes of a doctor.

REFERENCES

1. Amit Saxena, Leeladhar Kumar Gavel and Madan Madhaw Shrivas, "Rough Sets for Feature Selection and Classification: An Overview with Applications", *International Journal of Recent Technology and Engineering (IJRTE)*, 2014; **3**(5).
2. Bichen Zheng, Sang Won Yoon, arah S. Lam, " Breast cancer diagnosis based on feature extraction using a hybrid of K-means and support vector machine algorithms", Elseiver, *Expert System with application,*, 2014; 1476-1482.
3. Hesham Arafat, Sherif Barakat and Amal F. Goweda, " Using Intelligent Techniques for Breast Cancer Classification", *International Journal of Emerging Trends & Technology in Computer Science (IJETTCS)*, 2012; **1**(3).
4. Lynne E. Parker, " Notes on Multilayer, Feedforward neural networks CS494/594", projects in Machine Learning Spring, 2006,

website: http://web.eecs.utk.edu/~leparker/Courses/CS594-spring06/handouts/Neural-net-notes.pdf.

5. P. K. Nizar Banu and H. Hannah Inbarani, "Performance Evaluation of Hybridized Rough Set based Unsupervised Approaches for Gene Selection, *International Journal of Computational Intelligence and Informatics*, 2012; **2**(2).
6. Niloofar Mohammadzadeh, Reza Safdari, and Farshid Mohammadzadeh, "Using Intelligent Data Analysis in Cancer Care: Benefits and Challenges", *Journal of Health Informatics in Developing Countries*, 2014; **8**(2).
7. Parag Deoskari, Dr Divakar Singh, Dr. Anju Singh, "An Efficient Support Based Ant Colony Optimization Technique for Lung Cancer Data", *International Journal of Advanced Research in Computer and Communication Engineering*, 2013; **2**(9).
8. P.Ramachandran, N.Girija and T.Bhuvaneswari, "Early Detection and Prevention of Cancer using Data Mining Techniques", *International Journal of Computer Applications*, 2014; **97**(13).
9. K.K. Ramesh Kumar, A.Anbumani, "Medical Image Segmentation using Multifractual Analysis", *International Journal of Inventions in Computer Science and Engineering*,2014; **1**(3).
10. Syed Imran Ali and Waseem Shahzad, " Feature Subset Selection Method based on Conditional Mutual Information and Ant colony Optimization", *International Journal of Computer Applications*, 2012; **60**(11).
11. Selva Mary. G, Likhesh N, Kolhe, Kanchan D. Patil, " Classification of Micro Array Gene Expression using kNN, SVM and Naïve Classifiers", *International Journal of Scientific and Research Publications*, 2014; **4**(11).
12. P.S. Shelokar, Patrick Siarry, V.K.Jayaraman, B.D. Kulkarni, " Particle Swarm and Ant Colony algorithms Hybridized for improved Continues Optimization", *Applied Mathematics and Computation*, Elsevier, 2007; **188**(1): pages 129-142.
13. Tuan Zea Tan, Chai Quek, Geok See Ng, Khalil Razvi, " Ovarian Cancer diagnosis with complementary learning fuzzy neural network", Elseiver, *Artificial Intelligence System*, 2008; **43**: pages 207-222.
14. Dr.K. Usha Rani, " Parallel Approach for Diagnosis of Breast Cancer Using Neural Network Technique", *International Journal of Computer Application*, 2010; **10**(3).
15. Di Wang , Chai Quek, and Geok See Ng, " Ovarian cancer diagnosis using a hybrid intelligent system with simple yet convincing rules", Elseiver, 2014; pages 15.
16. Zakaria Suliman Zubi, Rema Asheibani Saad, "Improves Treatment Programs of Lung Cancer Using Data Mining Techniques", *Journal of Software Engineering and Applications*, 2014; **7**: pages 66-77.
17. http://www.nuh.com.sg/#.