

A FRGSNN Hybrid Feature Selection Combining FRGS filter and GSNN wrapper

Bichitrananda Patra

*Associate Professor, Computer Science & Engineering
SOA University, Bhubaneswar, Odisha, India*

Sujata Dash

*Professor, Department of Computer Science
North Orissa University, Baripada, Odisha, India.*

Abstract—How to selecting a small subset out of the thousands of genes in microarray data is important for accurate classification of phenotypes. Widely used methods typically rank genes according to their differential expressions among phenotypes and pick the top-ranked genes. While microarrays can measure the levels of thousands of genes per sample, case-control microarray studies usually involve no more than several dozen samples. Standard classifiers do not work well in these situations where the number of features (gene expression levels measured in these microarrays) far exceeds the number of samples. Selecting only the features that are most relevant for discriminating between the two categories can help construct better classifiers, in terms of both accuracy and efficiency. We observe that feature sets so obtained have certain redundancy and study methods to minimize it. We propose the minimum redundancy and maximum relevance feature selection framework. In this paper we have applied two general approaches of feature subset selection, more specifically, wrapper and filter approaches and then created a new model called hybrid model by combining the characteristics of the two specified models for gene selection. We have also compared the gene selection performance of the filter model, wrapper model and hybrid model. This lead to significantly improved class predictions in extensive experiments on 4 gene expression data sets: CNS, Leukemia, Lung and Brain Tumor. Improvements are observed consistently among 3 neural network algorithms classification methods such as Linear Vector Quantization (LVQ), Self-Organization Map (SOM) and Back Propagation (BP).

Keywords - Fuzzy rough sets, Evolutionary algorithms, Adaptive Neural-Network, Feature selection.

I. INTRODUCTION

The selection of efficient feature extraction techniques and predictive models provide high classification accuracy for microarray dataset. As such, the investigations show that a small number of gene expression data has strong correlation with certain phenotypes compared to the total number of genes available. Consequently, selections of differentially expressed relevant predictor genes correctly analyse gene expression profiles and also play a crucial role in classification process. The subset of potential genes identified by feature selection technique correctly distinguish the sample classes. Therefore, a good selection method for genes, relevant for sample classification—based on the number of genes investigated—is needed to increase the predictive accuracy and to avoid incomprehensibility.

An application of the gene expression data analysis is cancer classification. Usually, expression levels of the genes in microarray dataset are expressed in the form of a matrix, where each column corresponds to the expression level of one gene and each row to one sample. Each cell of this matrix underlies the expression level of a gene under a specific condition, and is represented by a real value. Typically these datasets contain thousands of genes while the number of conditions (tissue samples) ranges from tens to hundreds. To cope with this kind of extremely high-dimensionality, traditional machine learning methods such as linear discriminate analysis [12] and nearest- neighbour [18] and more advanced methods such as fuzzy logic [10] and neural networks [9] cannot be employed effectively and efficiently.

Feature selection [24] is a pre-processing technique whose main objective is to reduce the original training set. By removing noisy and irrelevant data—harmful for the majority of machine learning methods—data reduction can help to avoid the requirement of excessive storage and time, easing and enabling machine learning techniques to deal with large data sets. The best known data reduction processes are feature selection (FS) [11], feature generation/extraction [6], attribute discretization and prototype generation [23].

One of the most successful families of data reduction methods has been originated by evolutionary computation [2, 4]. Evolutionary algorithms are search algorithms inspired by natural genetics to evolve solutions, have been applied to different data reduction problems, modelling them as combinatorial problems [14]. However an evolutionary algorithm such as, genetic algorithm (GA) is initiated with a set of feasible solutions (chromosomes) called the population. Evaluation of each of the population is based on its fitness value computed from fitness

function i.e., more strong the solution, higher is the possibility to reproduce. This process is repeated until the stopping condition (for example, number of populations or quality of the solution) is satisfied. As a result, Genetic Algorithm searches all feasible solution regions which improve the chance of searching unexplored regions and possibility of achieving an overall optimal/near optimal solution. A remarkable number of evolutionary data reduction techniques have been focused on optimizing the adaptive Neural Network learning rule.

Fuzzy sets [26] and rough sets [25] address two important, complementary characteristics of imperfect data and knowledge: the former model express vague information of the objects belong to a set or a relation to a given degree, while the latter provide approximations of concepts in the presence of incomplete information [17,16]. A hybrid fuzzy rough set model was first proposed in and later extended and/or modified by many authors, being applied successfully in various domains [27].

On the other hand, much research on neural networks has been conducted. The neural networks can construct nonlinear decision boundaries without prior assumptions about the statistics of input data and have been used in many feature identification tasks including pattern recognition [13], image segmentation, and model free gait recognition method used as a biometric [8]. In particular, they have strong classification and learning power to represent implicit knowledge of the given data.

In this paper, a hybrid model (FRGSNN) for data reduction combining fuzzy rough (FR) sets and an evolutionary genetic search algorithm (GS) is used. The feature selection (FS) is performed by fuzzy-rough and genetic algorithm in two different approaches, filter and wrapper searches [15]. The wrapper search (GS-NN) uses the machine learning algorithm Nearest Neighbour (NN) to evaluate GS solutions [14]. The filter search (FRGS) compute the core reduct set measuring the quality of the features in fuzzy rough lower approximation using weak gamma method which calculates the discriminative features that must appear in every valid reduct set. The computed reduct sets are then used by the evolutionary search method (GS) as a set of candidate solutions for the considered problem. The goodness of each candidate solution is evaluated based on its fitness, and the population evolves by selection, crossover and mutation. The selected individuals are used to search for a global optimal subset of features. The rest of this paper is organized as follows.

We begin with a brief overview of the Rough set, Fuzzy Rough set, machine learning algorithms and proposed learning scheme of evolutionary hybrid search method are described in Section 2 and in Section 3 classification algorithms are introduced. Section 4 illustrates the four microarray datasets. The experimental framework and results are described in Section 5. Finally, the conclusion and future work are presented in Section 6.

II. OVERVIEW OF FEATURE SELECTION TECHNIQUES

2.1. *Rough Set Attribute Selection*

In microarray gene expression data analysis, data is available in the form of a collection of real-valued vectors. The dimension of these vectors is very high which causes problem in identifying marker genes causing cancer. Thus a technique that can reduce the dimensionality using information contained in the dataset and preserves the meaning of the attributes (semantic preserving) is clearly desirable. However semantic-destroying dimensionality reduction techniques transform data irreversibly, whereas semantic-preserving dimensionality reduction techniques widely known as attribute selection, attempt to retain the meaning of the original attribute set.

Rough set theory can be used to discover data dependencies (redundant data) and reduction of the attributes contained in the dataset without using any other information except the data available in the dataset [25]. The main shortcoming of rough set theory is that it cannot deal with real-valued problems, whereas many real world problems are real-valued. Fuzzy-rough set theory is a mathematical technique which is capable of reducing crisp and real valued attribute datasets.

2.2. *Fuzzy-Rough attribute selection*

The core of the Fuzzy-Rough Attribute Selection is the concept of indiscernibility relation which partitions the domain. Given a set of attributes as the objects of the domain, objects with the same attribute values are indiscernible and would belong to the same block of the partition. The task is to approximate a rough (imprecise) concept in the domain by a pair of exact concepts. These two exact concepts are the lower and upper approximations which are determined by the indiscernibility relation. In figure 1, each cell represents an equivalence class which is obtained from the indiscernibility relation. The lower approximation is a set of objects definitely belonging to the rough concept, whereas the upper approximation is a set of objects possibly belonging to the same [15]. Therefore, these two can be constructed from equivalence classes.

2.3. *Fuzzy-rough set attribute selection*

2.3.1. *Fuzzy Equivalence Classes*

Gene expression data is a real-valued dataset, thus we employ fuzzy-rough set for attribute selection. Fuzzy equivalence classes are the core of fuzzy-rough set approach. In this case decision values and conditional values should all be fuzzy. Fuzzy S-lower and S-upper approximations are defined as:

$$\mu_{\underline{S}X}(x) = \sup_{F \in \mathcal{S}} \min(\mu_F(x), \inf_{y \in U} \max\{1 - \mu_F(y), \mu_X(y)\}) \quad (1)$$

$$\mu_{\overline{S}X}(x) = \sup_{F \in \mathcal{S}} \min(\mu_F(x), \inf_{y \in U} \min\{1 - \mu_F(y), \mu_X(y)\}) \quad (2)$$

where S is an equivalence class, X is the concept to be approximated and F is a fuzzy equivalence class belonging to U/S.

2.3.2. Fuzzy-rough reduction process

Based on notion of fuzzy lower approximation, for real-valued attributes, attribute selection in fuzzy-rough set is achieved. The membership degree of an observation of the universe x, belonging to fuzzy positive region, can be defined as:

$$\mu_{POS_S}(Q) = \sup_{x \in U/S} \mu_{\underline{S}X}(x) \quad (3)$$

Using fuzzy positive region the fuzzy dependency function is defined as:

$$\gamma_S(Q) = \frac{|\mu_{POS_S}(Q)(x)|}{|U|} = \frac{\sum_{x \in U} \mu_{POS_S}(Q)(x)}{|U|} \quad (4)$$

The dependency of O on S is a proportion of observations that are discernible out of the whole dataset.

2.4. Genetic Algorithms

Genetic algorithms are derivative-free optimization method motivated by biological phenomenon of natural selection and evolution and have wide applications in many different problems [2,4]. It generates new population of chromosomes that is considered as the candidate solution in each iteration. Each chromosome is a string of encoded binary, real, etc., version of a candidate solution. Every string is associated with a fitness measure evaluated by an evaluation function that indicates its fitness for the problem. Then the standard genetic operators such as selection, crossover and mutation are applied to the initial random population to compute a new generation of individuals. The individuals having more fitness value are likely to be selected for the next generation. The variation is introduced to the selected individuals for obtaining global optimum solution using crossover and mutation [7].

2.5. K-nearest neighbour

The k-nearest neighbour is a popular nonparametric supervised learning algorithm. This algorithm is also called instance based learning method. It classifies a new sample using majority of the k-nearest neighbour and the new sample is assigned the class that is most frequent among these k-neighbours [8]. K-nearest neighbour is simple, easy to implement, handles large amount of data and indifferent to noisy data. In this work, the neighbours are calculated using Euclidian distance and then this algorithm evaluates the significance of feature subset using 10-fold cross validation. No user specified parameters are 1-NN classifier.

2.6. The Proposed Learning Scheme - Evolutionary hybrid fuzzy-rough-genetic search method for feature selection

The first key element of the hybrid FRGSNN model is its fuzzy-rough feature selection method which computes the reduct sets. The second key element of FRGSNN model is its search method for selecting optimal subsets of features. To perform this job, we have chosen genetic search (GS) an evolutionary algorithm to perform the search. The reduct set produced from fuzzy rough reduction process is being used as the population of the genetic search in which only a reduced set of offspring is produced in each generation, generally two, in most of the cases. Parents are chosen to reproduce children and then a decision is being made to select the candidates from the population which will be deleted in order to accommodate the new candidates. The main task of the method is to search for subsets of features which increase the accuracy of the classifier. However, a second task should be to reduce the size of the subsets selected, provided it should not reduce the accuracy rate of the problem. The configuration details of the Genetic search (GS) is as follows:

- Codification: The Genetic search (GS) will encode each point in a solution space into a binary bit string called a chromosome. Each bit will represent the state of each feature in the reduct set derived from rough fuzzy reduction process (1 if the feature is selected; 0 if it is deleted).
- Crossover: A two-point crossover i.e., two crossover points are selected and the part of the chromosome string between these two points is then swapped to generate two children. In each generation, this operator is applied twice, obtaining two offspring.

- Mutation: Crossover exploits current gene potentials, but if the population does not contain all the encoded information needed to solve a particular problem, no amount of gene mixing can produce a satisfactory solution. For this reason, a mutation operator capable of spontaneously generating new chromosomes is included. The flip-bit mutation operator (changing the value of the selected allele from 0 to 1 and vice versa) is applied to each offspring produced, with a probability equal to a very low mutation rate.
- Selection of parents: The selection operation determines which parents participate in producing offspring for the next generation, and it is analogous to survival of the fittest in natural selection. Usually members are selected for mating with a selection probability proportional to their fitness values. A binary tournament procedure will be used to select parents in each generation.

III. CLASSIFICATION ALGORITHMS

3.1. LVQ Classifications

LVQ was applied successfully to areas such as audio compression, data compression, data transmission, facial recognition, radar signal processing, finance and insurance, production control, sale and marketing, and so on. Keeping all these issues in view, LVQ could be applied to such simple structured data, with higher confidence than that of other classification algorithm. One of the most amazing features of LVQ algorithm is that it can take very few vectors to obtain excellent classification results. The idea behind LVQ is to take away codebook vectors from the decision surfaces to clearly demarcate the class borders [22].

Assume that a number of 'codebook vectors' m_i (free parameter vectors) are placed into the input space to approximate various domains of the input vector x by their quantized values. Usually several codebook vectors are assigned to each class of x values, and x is then decided to belong to the same class to which the nearest m_i belongs.

$$\text{Let } c = \arg \min (||x - m_i||) \quad (5)$$

define the nearest m_i to x , denoted by m_c .

Values for the m_i that approximately minimize the misclassification errors in the above nearest-neighbor classification can be found as asymptotic values in the following learning process. Let $x(t)$ be a sample of input and let the $m_i(t)$ represent sequences of the m_i in the discrete-time domain. Starting with properly defined initial values, the following equations define the basic LVQ1 process [21]:

$$\begin{aligned} m_c(t+1) &= m_c(t) + \alpha(t)[x(t) - m_c(t)] \\ &\quad \text{if } x \text{ and } m_c \text{ belong to the same class,} \\ m_c(t+1) &= m_c(t) - \alpha(t)[x(t) - m_c(t)] \\ &\quad \text{if } x \text{ and } m_c \text{ belong to different classes,} \\ m_i(t+1) &= m_i(t) \text{ for } i \text{ not in } c. \end{aligned}$$

Here $0 < \alpha(t) < 1$, and $\alpha(t)$ may be constant or decrease monotonically with time. In the above basic LVQ1 it is recommended that α should initially be smaller than 0.1; linear decrease in time is used.

3.2. The Self-Organizing Map

T. Kohonen introduced the Self-Organizing Map (SOM) [20]. It is an unsupervised learning process, which learns the distribution of a set of patterns without any class information. It has the property of topology preservation. There is a competition among the neurons to be activated or fired. The result is that only one neuron that wins the competition is fired and is called winner-takes all neuron. SOMs may be one-dimensional, two-dimensional or multidimensional, but the most common ones are either one-dimensional or two-dimensional maps.

The number of input connections depends on the number of attributes to be used in the classification. The neuron with weights closest to the input data vector is declared the winner during the training. Then the weights of all of the neurons in the neighborhood of the winning neuron are adjusted by an amount inversely proportional to the distance. It clusters and classifies the data set based on the set of attributes used. The algorithm is summarized as follows [3]:

Step 1- Initialization:

Choose random values for the initial weight vectors $w_j(0)$, the weight vectors being different for $j = 1, 2, \dots, l$ where l is the total number of neurons

Step 2- Sampling:

Draw a sample x from the input space with a certain probability.

Step 3- Similarity Matching:

Find the best matching (winning) neuron $i(x)$ at time steps n by using the minimum distance Euclidean criterion

Step 4- Updating:

Adjust the synaptic weight vector of all neurons by using the update formula where, $\eta(n)$ is the learning rate parameter, and $h_{j, i(x)(n)}$ is the neighborhood function centered around the winning neuron $i(x)$. Both $\eta(n)$ and $h_{j, i(x)(n)}$ are varied dynamically during learning for best results.

Step 5- Continue with Step-2 until no noticeable changes in the feature map are observed.

3.3. Back Propagation Algorithm

The back propagation algorithm trains a given feed-forward multilayer neural network for a given set of input patterns with known classifications. When each entry of the sample set is presented to the network, the network examines its output response to the sample input pattern. The output response is then compared to the known and desired output and the error value is calculated. Based on the error, the connection weights are adjusted. The back propagation algorithm is based on *Widrow-Hoff delta learning rule* in which the weight adjustment is done through *mean square error* of the output response to the sample input "Vel98 Simulation Program". Master's book, The University of Texas at El Paso (Vel98 Velasquez, Guillermo. "A Distributed Approach to a Neural Network, El Paso, TX, 1998.) The set of these sample patterns is repeatedly presented to the network until the error value is minimized.

There are $L+1$ layers of neurons, and L layers of synaptic weights. We'd like to change the weights W and biases b so that the actual output x^L becomes closer to the desired output d .

Algorithm 3.9 Back propagation Algorithm

1. Forward pass. The input vector x^0 is transformed into the output vector x^L , by evaluating the equation

$$x_i^l = f(u_i^l) = f\left(\sum_{j=1}^{n_{l-1}} W_{ij}^l x_j^{l-1} + b_i^l\right) \quad (6)$$

from $l = 1$ to L .

2. Error computation. The difference between the desired output d and actual output x^L is computed.

$$\delta_i^L = f'(u_i^L)(d_i - x_i^L) \quad (7)$$

3. Backward pass. The error signal at the output units is propagated backwards through the entire network, by evaluating

$$\delta_i^{l-1} = f'(u_i^{l-1}) \sum_{j=1}^{n_l} \delta_j^l W_{ij}^l \quad (8)$$

from $l = L$ to 1 .

4. Learning updates. The synaptic weights and biases are updated using the results of the forward and backward passes,

$$\Delta W_{ij}^l = \eta \delta_i^l x_j^{l-1} \quad (9)$$

$$\Delta b_i^l = \eta \delta_i^l \quad (10)$$

These are evaluated for $l = 1$ to L . The order of evaluation doesn't matter.

IV. DATASETS ILLUSTRATION

A large amount of gene expression microarray data has become available in several datasets. In this paper, four public microarray datasets were used to assess the performance of the proposed classifier. First two dataset of Table 1 is collected from <http://datam.i2r.a-star.edu.sg/datasets/krbd/> (CNS embryonal tumour outcome, Lung Harvard) and last two from <http://www.gems-system.org/> (Leukaemia, Brain Tumour). The following is a brief description of these datasets.

Table 1: Gene expression datasets used in this paper

Dataset	Number of genes	Training Samples	Test samples	References
CNS	7129	60 (21:39)	0	Scott L. Pomeroy, et al. (2002)
ALL_AML_GEMS	7129	38 (27:11)	34(20:14)	Golub, et al. (1999)
Lung Harvard	12600	122 (58:21:20:6:17)	81(81:0:0:0:0)	A. Bhattacharjee, et al.(2001)
Brain_Tumor1_GEMS	5920	90 (60:10:10:4:6)	0	Pomeroy SL et al. (2002)

V. EXPERIMENTAL RESULTS AND DISCUSSION

In this paper, we have used 4 microarray datasets to ascertain the validity of the projected scheme. Out of 4 datasets, 2 datasets i.e., Central system (CNS) and ALL-AML are unit of binary category and also the remaining 2 i.e., Lung-Harvard and Brain-Tumor-GEMS are unit of multi-class datasets. All the learning algorithms used in this work are taken from WEKA machine learning package.

The experimental method is composed of feature selection and classification method. In the feature selection process, we have used fuzzy rough with genetic search as filter method (FR-GS), evolutionary wrapper method (GS-NN) and a hybrid method (FR-GS-NN) to select potential genes from the datasets. These methods reduce the computational complexity of the problem and select the most informative subset of genes from each dataset. Table 2 summarizes the subsets of features selected by FR-GS, GS-NN and FR-GS-NN from each microarray dataset. Then the representative subset of genes obtained from the above processes is classified through the individual adaptive neural network classifiers such as LVQ, SOM and BP. The test dataset of ALL-AML and Lung cancer are used by the learning algorithm for the purpose of prediction. For the remaining datasets, ten-fold cross validation was applied on training sets to measure the statistically reliable predictive accuracy.

Table 2 Selection of Attributes using FR-GS filter, GS-NN wrapper and FR-GS-NN hybrid method

Datasets	No. of Attributes before filtering	Selection of Attributes after filtering FRGS	Selection of Attributes after using GSNN	Selection of Attributes after using hybrid FRGSNN	Percentage of attributes selected by FRGSNN
CNS	7129	162	2814	45	0.63
ALL-AML	7129	162	3442	58	0.81
Lung Harvard	12600	234	3313	121	0.96
Brain-Tumor	5920	267	2111	108	1.82

From the results of 4 trained datasets, we can observe that the classifiers attain the best results when they work on particular datasets. However, there is no generalized strategy designed for tumour classification problems based on a wide range of different datasets so far.

For all component adaptive neural network classifiers, we have set the parameters of BP with 10 hidden nodes in a single hidden layer, 2 output nodes for binary datasets, learning rate varies from 0.01~0.10, momentum 0.9 and no of maximum iterations 500. In LVQ learning we have used, 25 codebook vectors, 0.01~0.10 learning rate, static learning function, momentum 0.1 and maximum iterations 1000. In SOM, we have used initial 5 X 5 map which has rectangle shape, neighbourhood size 16, static learning function, learning rate ranges from 0.1~0.9 and maximum iterations is set to 2000.

VI. RESULTS

In Table 3 the classification accuracy of LVQ, SOM and BP with kappa statistics computed for the subset of features obtained from filter (FRGS), wrapper (GSNN) and hybrid (FRGSNN) for all four datasets are shown. Out of these 4 datasets, first two belong to binary class problem and last two belong to multi-class problem.

In the entire three feature reduction model, when compared the predictive accuracy of individual classifiers, it is observed that the performance of Self Organizing Map (SOM) is not so significant with respect to other classifiers such as Linear Vector Quantization (LVQ) and Back Propagation (BP) algorithm for all the datasets. The least and highest predictive accuracy evaluated by SOM is 59.7222% and 79.3842% for binary class ALL—AML and multi-class problem Lung dataset respectively. The least accuracy is obtained from filter model whereas highest is obtained from both wrapper and hybrid model for SOM.

Similarly, by analysing the performance of predictive accuracy of LVQ neural network in Table 4 we find, the least 68.33% is achieved by CNS (binary class) and the highest 93.4544% is by ALL-AML (binary class). The least and highest accuracy are obtained from FRGS filter model and FRGSNN hybrid model respectively.

Further the analysis of the result of component Back Propagation (BP) neural network in Table 4 shows that, minimum prediction accuracy of 68.4729% is given by Lung (multi class) for FRGS filter model. BP has registered maximum classification accuracy of 85.33% for CNS (binary-class) in FRGSNN hybrid model.

Table 4 summarizes the classification accuracy of kappa statistic for all the subsets obtained by 3 different feature selection algorithms for all the datasets. As can be observed from the table, adaptive neural network classifier performs significantly better than the individual base learners. The classification accuracy and the value of kappa

statistic (k) obtained from the proposed method has established a proper balance between accuracy and diversity. It is worth to note that, a small value of k denotes high diversity among the learners. Our method has achieved proper balance between diversity and accuracy and at the same time fulfilled the objective of finding a generalized prediction model.

Table 3: Best estimated accuracy results using three elementary neural network classifiers

Datasets	FR-GS			GS-NN			FR-GS-NN		
	LVQ	SOM	Back Propagation	LVQ	SOM	Back Propagation	LVQ	SOM	Back Propagation
CNS	68.33	66.6667	81.6667	68.33	66.6667	84.33	68.33	69.67	85.33
ALL_AML	73.6111	59.7222	69.4444	90.8333	68.0556	70.8333	93.4544	70.8333	76.3889
Lung Harvard	86.6995	77.8325	68.4729	89.6552	79.3842	80.3103	90.1478	79.3842	81.4729
Brain_Tumor1	74.444	66.6667	68.8889	78.8889	68.3333	70	83.3333	71.1111	71.6667

Table 4 : Results from Kappa statistic

Datasets	FR-GS Accuracy of Kappa statistic (k)	GS-NN Accuracy of Kappa statistic (k)	FR-GS-NN Accuracy of Kappa statistic (k)
CNS	0.41	0.92	0.99
ALL_AML	0.58	0.84	0.92
Lung Harvard	0.76	0.84	0.92
Brain_Tumor1	0.72	0.74	0.89

VII. DISCUSSION

Feature selection plays an important role in enhancing the classification accuracy of the problem. Potential features influence the predictive ability of the classification model. Hence, efficient feature selection method needs to be employed to select optimal features for the problem which results in increasing the classification accuracy. In this study, three types of feature selection methods classifier are introduced for classifying gene expression datasets. The proposed method outperforms the results published recently. We found the discrimination ability of the LVQ classifier is highly significant when search strategy like filter, wrapper and hybrid methods are concerned. Hence, these significant improvements of the proposed scheme can be attributed to the fact that the hybrid feature selection mechanism identifies different combinations of genes that enhances the classification accuracy there by improves the stability and generalization capability.

In this problem, we emphasized both feature selection and classification aspects of the problem. But we have not studied the importance of individual gene and relation among them which specifically help to get more information about the disease. This aspect of the problem is left for our future research.

The only shortcoming of this method is its high computational complexity which increases the cost of the problem. However, the remarkable achievement in classification accuracy can compromise with the cost happened due to wrong diagnosis or prognosis of the disease.

VIII. CONCLUSION

Experimental results show that the evolutionary feature selection techniques effectively selected subset of marker genes for each component learner and increased the discriminatory power of the neural network classifier. Also it is evident from the study that LVQ classifier performs better than SOM and BP classifiers for all datasets. Identification of potential genes from microarray dataset for classification problems has been a challenging task for the researchers. The proposed method has achieved the highest averaged generalization ability compared to its counterparts and established an acceptable level of diversity among the base learners for majority of the analyzed benchmark datasets. We believe, our suggested method has addressed the challenge to a great extent.

REFERENCES

- [1] ArindamBhattacharjee, et al., Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses, PNAS, The National Academy of Sciences, USA, vol. 98, no. 24, 2001, 13790-13795.

- [2] Eiben AE, Smith JE; Introduction to Evolutionary Computing. Natural Computing, Springer-Verlag, Berlin, 2003.
- [3] Face Recognition Using Self-Organizing Map and Principal Component Analysis, Dinesh Kumar, C.S.Rai and Shakti Kumar, 2007.
- [4] Ghosh A, Jain LC (eds) Evolutionary computation in data mining. Springer-Verlag, Berlin, 2005.
- [5] Golub, et al., Molecular classification of cancer: class discovery and class prediction by gene expression monitoring, 1, Science, vol. 286, 1999:531-537.
- [6] Guyon I et al.; (eds) Feature Extraction: foundations and applications. Springer, Berlin, 2006.
- [7] Heesung Lee et al.; A new genetic feature selection with neural network ensemble, International Journal of Computer mathematics, vol.86, No. 7, July 2009, 1105-1117.
- [8] Heesung Lee et al.; Neural Network Ensemble with probabilistic fusion and its application to gait recognition, Neuro computing 72,2009; 1557-1564.
- [9] J. Han et al., Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks, Nature Med. 7;2001, 673-679.
- [10] L. Ohno-Machado et al.; Classification of gene expression data using fuzzy logic, J. Intell. Fuzzy Syst. 12;2002, 19-24.
- [11] Liu H, Motoda H (eds) Computational methods of feature selection. Chapman & Hall/Crc data mining and knowledge discovery series, Chapman & Hall/Crc, London, 2007.
- [12] M. Xiong et al.; Feature (gene) selection in gene expression-based tumor classification, Molecular Genetics and Metabolism 73;2001, 239-247.
- [13] N. Karayiannis and Y. Xiong, Training reformulated radial basis function neural networks capable of identifying uncertainty in data classification, IEEE Trans. Neural Networks 17; 2006, pp. 1222-1234.
- [14] Pappa GL, Freitas AA; Automating the design of data mining algorithms: an evolutionary computation approach. Natural computing. Springer, Berlin, 2009.
- [15] R. Jensen and Q. Shen, Fuzzy-rough attributes reduction with application to web categorization, Fuzzy Sets Syst.141;2004, 469-485.
- [16] S. Dash, B.N. Patra; Redundant Gene Selection based on Genetic and Quick-Reduct Algorithms", IJMIA: International Journal on Data Mining and Intelligent Information Technology Application, ISSN: 2234-3660, 15th Feb 2013. (Accepted and in press).
- [17] S. Dash, B.N. Patra; Rough Set Aided Gene Selection for Cancer Classification, published in IEEEExplore 7th International Conference on Computing and Convergence Technology (ICCCT), 2012 on 3-5 Dec.,2012. Page(s):290-294, Print ISBN : 978-1-4673-0894-6.
- [18] S. Dudoit et al.; Comparison of discrimination methods for the classification of tumors using gene expression data, Technical report, University of California, Berkeley, CA, 2000.
- [19] Scott L. Pomeroy, et al., Prediction of central nervous system embryonal tumour outcome based on gene expression, Nature, vol. 415,2002, 436-442.
- [20] Self Organizing Maps, Tom Germano, March 23, 1999.
- [21] T.V. Prasad, Maitrei Kohli, "Vector Quantization of Microarray Gene Expression Data", Proceedings of the World Congress on Engineering 2010 Vol I WCE 2010, June 30 - July 2, 2010, London, U.K.
- [22] The Learning Vector Quantization Program Package Version 3.1 (April 7, 1995) Prepared by the LVQ Programming Team of the Helsinki University of Technology Laboratory of Computer and Information Science Rakentajanaukio 2 C, SF-02150 Espoo FINLAND.
- [23] Triguero I et al.; A taxonomy and experimental study on prototype generation for nearest neighbour classification. IEEE Trans Syst Man Cybern Part C Appl Rev 42(1),2012:86-100.
- [24] Yang C. Huei et al.; IG-GA: A hybrid Filter/Wrapper Method for Feature Selection of Microarray Data. Journal of Medical and Biology Engineering, 30(1), 2009:23-28.
- [25] Z. Pawlak, Rough sets, Int. J. Inform. Comput. Sci. 11;1982, 341-356.
- [26] ZadehLA Fuzzy sets. Inf Control 8(3),1965:338-353.
- [27] Zhai J; Fuzzy decision tree based on fuzzy-rough technique. Soft Compute 15(6); 2011:1087-1096.