Journal of Engineering Sciences

Vol 11, Issue 4 , April/2020
ISSN NO:0377-9254

HYBRID FEATURE SELECTION OF CORRELATION COEFFICIENT WITH PSO ON MICRO ARRAY GENE EXPRESSION DATA

G. Chandini Alekhya[1], D. Sruthi[2] ,J. Abhilasha[3] , K.Vineetha[4],V.Lakshman Narayana[5]

[1,2,3,4]IV B.Tech, Department of Information Technology, Vignan's Nirula Institute of Technology & Science for Women, Peda Palakaluru, Guntur-522009, Andhra Pradesh, India.

[5]Assoc.Professor & HOD, Department of Information Technology, Vignan's Nirula Institute of Technology & Science for Women, Peda Palakaluru, Guntur-522009, Andhra Pradesh, India.

lakshmanv58@gmail.com

ABSTRACT

Diagnosis of cancer is one of the most emerging clinical applications in microarray gene expression data. However, cancer classification on microarray gene expression data still remains a difficult problem. The main reason for this is the significantly large number of genes present relatively compared to the number of available training samples. In this paper, we propose a hybrid feature selection approach that combines the correlation coefficient with particle swarm optimization. Gene expression data is widely used in disease analysis and cancer diagnosis. In this study, we applied both the information gain and correlation-based feature selection method as filter approaches, and an improved binary particle swarm optimization as a wrapper approach to implement feature selection; selected gene subsets were used to evaluate the performance of classification. Experimental results show that by employing the proposed method fewer gene subsets needed to be selected and better classification accuracy could be obtained. Diagnosis of cancer is one of the most emerging clinical applications in microarray gene expression data. However, cancer classification on microarray gene expression data still remains a difficult problem. The main reason for this is the significantly large number of genes present relatively compared to the number of available training samples. In this work, a hybrid feature selection approach that combines the correlation coefficient with particle swarm optimization is proposed. After the process of feature selection is performed, the selected genes are subjected to Extreme Machine Learning Classifier. Experimental results show that the proposed hybrid approach reduces the number of effective levels of gene expression and obtains higher classification accuracy and uses fewer features compared to the same experiment performed using the traditional tree-based classifiers like J48, random forest, random trees, decision stump and genetic algorithm as well.

Keywords: Feature selection, Correlation coefficient, Particle Swarm Optimization, Extreme Learning Machine, Gene expression data, microarray.

1. INTRODUCTION

DNA microarray technology allows simultaneous monitoring and measuring of thousands of gene expression activation levels in a single experiment[1]. This technology is currently used in medical diagnosis and gene analysis. Many microarray research projects focus on clustering analysis and classification accuracy. In clustering analysis, the purpose of clustering is to analyze the gene groups that show a correlated pattern of the gene expression data and provide insight into gene interactions and function. Research on classification accuracy is aimed at building an efficient model for predicting the class membership of data, produce a correct label on training data, and predict the label for any unknown data correctly[2]. Typically, gene expression data possess a high dimension and a small sample size, which makes testing and training of general classification methods difficult. In general, only a relatively small number of gene expression

data out of the total number of genes investigated shows a significant correlation with a certain phenotype [41] [42]. In other words, even though thousands of genes are usually investigated, only a very small number of these genes show a correlation with the phenotype in question.

Over the few decades, bioinformatics has become a more and more notable research field since it allows biologists to make full use of the technologies in computer science and computational statistics to analyze the data of an organism at the genomic, transcriptomics and proteomic levels1,2,3[3]. One of the major tasks in biomedicine is the classification and the prediction of microarray data3. With the rapid development of DNA microarray technology, classification of microarray data is a challenging task since gene expression datasets are often with thousands of genes but a small number of samples4. Tumor classification is one of the conventional problems in microarray gene expression data, and includes tumor detection and prediction of some rare diseases5[4]. These studies are of tremendous importance for accurate cancer diagnosis and subtype recognition. Because of the limited availability of effective samples compared to thousands or even tens of thousands of genes in microarray data, many computational methods fail to identify a small portion of important genes, and it increases learning costs and deteriorates learning performance6,7. In general, cancer classification for microarray data involves data collection, preprocessing, gene selection, and so on. The goal of classification is to build efficient and effective gene selection methods, which reduce the dimensionality of microarray data to improve the classification accuracy of cancer gene expression datasets[5][6]. The aim of gene selection is to reduce the dimensionality of microarray data in order to enhance the accuracy of classification task8. Gene selection methods can reduce the number of irrelevant and noisy genes and select the most related genes to improve the classification results, which decrease the computational costs and improve the cancer classification performance9[7].

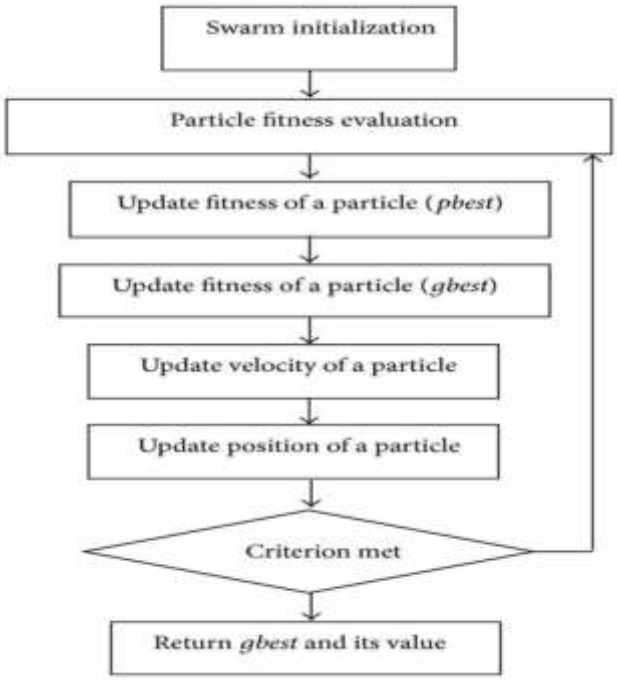


Fig1 Filter Model Approach

In the filter model approach a filtering process precedes the actual the classification process. For each feature a weight value is calculated, and features with better weight values are chosen to represent the original data set[8]. However, the filter approach does not account for interactions between features.

The wrapper model approach depends on feature addition or deletion to compose subset features, and uses evaluation function with a learning algorithm to estimate the subset features[9]. This kind of approach is similar to an optimal algorithm that searches for optimal results in a dimension space [43] [44]. The wrapper approach usually conducts a subset search with the optimal algorithm, and then a classification algorithm is used to evaluate the subset. Particle swarm optimization (PSO) is a population-based stochastic optimization technique, which was developed in 1995 [10].

PSO simulates the social behavior of organisms, such as birds in a flock or fish in a school, and can be described as an automatically evolving system. In PSO, each single candidate solution can be considered "an individual bird in the flock", that is, a particle in the search space. Each particle makes use of its own memory and knowledge gained by the swarm as a whole to find the best (optimal) solution. In 1997 a binary version of PSO (BPSO) [11] to solve discrete problems. In BPSO, each particle represents its position by either the binary value {0} or {1}, and the velocity is treated as a probability change of the particle position. However, BPSO has the same disadvantage as other evolutionary algorithms [45] [46]. After several generations, these algorithms tend to easy get trapped in a local optimum, which might prevent them from converging towards a global optimal solution. In order to circumvent the premature convergence at a local optimum, we incorporated a simple Boolean operation to create a new gBest position[12]. This new gBest replaced the original gBest, so that all particles were able to leave the local optimal.

Compared with the filter methods, the wrapper models select feature subsets with higher prediction accuracy with some search methods, and then their results are evaluated by a certain learning algorithm33[13]. The search methods in wrapper are divided into sequential strategies and random strategies [47] [48]. Recently, some wrapper-based approaches have been provided and widely applied in bioinformatics, such as genetic algorithm34 (GA), particle swarm optimization35 (PSO), Ant colony optimization36 (ACO), and so on. Although these approaches have obtained excellent performance in gene expression data analysis, some congenital drawbacks still puzzle themselves such as excessive computational cost of GA and local optimum of PSO36[14]. The ACO algorithm is inspired by the behavior of real ants. Since ACO needs no heuristic information for searching an optimal minimal subset every time, it is especially an attractive approach to feature selection37[15]. The ACO-based feature selection enables to efficiently balance between exploration and exploitation, and then can find more important features by taking advantage of the parameter adjustment and feature significance38. It has intelligent searching, global optimization, robustness and positive feedback; so many scholars have paid more attention to ACO39.

Statistical analysis of differentially expressed genes helps to assign them to different classes[16]. This process improves the understanding of basic biological processes in the system. Using the concept of technology of microarray gene expression, it is possible to study the simultaneous activity of thousands of genes. The relative abundance of mRNA in the gene can be found by using gene expression profiles [17]. Results obtained represent the state of the cell. Discriminant analysis of microarray data is an excellent tool for medical diagnostics of diseases, treatment and prevention[18].

The main purpose of the classification is to build an effective model that can identify differentially expressed genes and could also be used to identify classes in the unknown samples [19]. Some of the challenges in the microarray data are the smallest number of training and testing data available, the higher dimensionality of the data and the variations that could sneak in experiments performed to estimate the levels of gene expression. The two main tasks in the analysis of gene expression microarray are feature Selection and classification.

To perform the classification process with an acceptable level of accuracy, the process of feature selection becomes crucial. Microarray gene expression data contains hundreds of thousands of genes or feature information[20]. Only a small subset of genes exhibit strong correlation between them. Feature selection is a process that effectively selects differentially expressed genes in the dataset and

forms a new subset for efficient classification. There may be situations in which a low-ranked gene could perform well in the rankings and a critical gene could be left out in the selection of functions [21].

2. LITERATURE SURVEY

Narayana, V.L. [22] proposed a classification algorithm called ID3, which introduces the concept of information gain. Information gain is a measure based method, which is usually used to select best split attributes in decision tree classifiers. The measure indicates to what extent the entire data's entropy is reduced, and identifies the value of each specific attribute. Each feature basis obtains an information gain value, the amount of which is used to decide whether the feature is selected or deleted[23]. Therefore a threshold value for selecting a feature must first established; a feature is selected when the information gain value of this feature is bigger than the threshold value.

Relief algorithm is one of the widely applied filter-based feature selection models and has great classification efficiency[24]. In addition, this algorithm does not limit data types and can effectively deal with nominal or continuous features, missing data and noisy tolerance46. The principle of this algorithm is that the stronger correlation of classification makes the similar samples closer. On the contrary, the inhomogeneous samples are kept away.

ACO algorithm is one of the applications of wrapper-based feature selection methods and a probabilistic technique for solving computational problems to reduce the search path to find the optimal path through graphs, which can be usually used to find an optimum subset of features48[25]. The ACO algorithm has the strong robustness and the great performance on resolving the complex optimization problem, and is state-of-the-art for addressing the optimization problem of feature selection. It requires a problem that can describe a graph, where the nodes indicate features with edges among nodes and describe the next option of feature49. This optimal feature subset search is an ant path through graph where the minimum number of the visited nodes is suitable with the traversal stopping criterion48[26].

Previous studies have proposed hybrid models combining the advantage of filter and wrapper methods in a single approach [27]. These hybrid models employ filter method for initially screening out majority of irrelevant features and reducing the computational complexity of wrapper method. But this affects the performance of hybrid model constraining the wrapper method to reply only on those features selected by a single filter which may screen out some relevant biomarkers even from the chance to be considered in the wrapper evaluation [28]. Based on the assumption, there is no single filter that performs best for any given problem, this study considers the decision of ensemble of filters to address this problem.

3. PROPOSED METHOD

DNA Microarray technology (also called 'DNA chips') has become a powerful tool for biologists to monitor the gene expression levels within an organism [29]. This technology enables researchers to simultaneously measure the expression levels of a large number of genes. Gene expression data generally includes thousands of genes (high dimensionality), as well as a small number of samples[30]. It also contains numerous irrelevant and redundant features. Microarray technology is used most within medical fields, in order to learn about what causes diseases and how to treat them [31]. Researchers have figured out that the mutations in DNA may sometimes be the cause for certain diseases, like breast cancer. The mutation of certain known genes, is known as being the cause for some diseases. However, there is no one type of mutation that causes all diseases [32]. DNA Microarray data analysis is therefore used in order to discover and detect general mutations within DNA. Microarray gene expression data technology has had a massive effect on cancer research. It is a powerful technique when it comes to diagnosing and identifying the disease genes for human cancers [33]. Moreover, it has been vastly used to identified cancer-related genes using feature selection methods [34].

In gene expression Microarray data analysis, feature selection techniques are typically used to find the informative genes. Feature selection is how differentially expressed genes are discovered [35].

The process of feature selection is also called gene prioritisation, or biomarker discovery . Microarray data analysis process is challenging task, as there are ultimately too few samples, that in turn have too many features. The data sparsity of microarray exists due to the process of experiments. Many microarray data sets have missing values which affects the post-processing

PSO is an evolutionary optimization technique based on swarm intelligence developed by Kennedy and Eberhart in 1995 [36]. Unlike other evolutionary algorithms, PSO neither has complicated evolutionary operator nor has many parameter for tuning and solves the optimization problem exploiting the feeding characteristics of animals, birds and fish. In recent years, it has proven to solve diverse optimization problems with quick convergence rate and has attracted much attention of many researchers worldwide [37]. PSO algorithm starts with a population (called swarm) of random solutions. This initial population evolves iteratively to find optimal solution for the problem to be optimized. Each individual (called particle) in the population corresponds to a fitness value determined by the function to be optimized and has a velocity, which enables them to move through the search space. Thus, each particle represented by its position and velocity During the optimization, particles keep updating its position themselves using its previous position and its current velocity. The current velocity is determined using two cognitive aspects, individual learning (pBest) and learning from a social group (gBest) to move towards the global optimal solution of the problem. These principles are formulated as

$$x_i(t+1) = x_i(t) + v_i(t+1)$$

$$v_i(t+1) = \underbrace{w(t)v_i(t)}_{\text{inertial parameter}} + \underbrace{c_1 r_1 \left(pbest_i(t) - x_i(t)\right)}_{\text{particle best velocity component}} + \underbrace{c_2 r_2 \left(gbest(t) - x_i(t)\right)}_{\text{global best velocity component}}$$

Algorithm : PSO

1. Initialize the population S with solution space;

2. Repeat:

a. Evaluate fitness of each particle position (p) using objective function

b. If fitness(p) > pbest then pbest = fitness(p).

c. Set gbest= best of pbest

d. Update velocity and position of each particle

3. Until convergence is achieved or termination criteria are satisfied 4. Final gbest is optimal solution

```
def runEML():

srhl_tanh=MLPRandomLayer(n_hidden=6,activation
_func='tanh')

cls = GenELMClassifier(hidden_layer=srhl_tanh)

dimensions = X_train.shape[1]

 ps.discrete.BinaryPSO()

optimizer=ps.discrete.BinaryPSO(n_particles=20,dim
ensions=dimensions,options=options)cost, pos =
optimizer.optimize(fx.sphere, iters=10)

 X_selected_features = X_train[:,b==1]

 text.insert(END,"Total features after applying PSO :
"+str(len(X_selected_features))+"\n\n")

cls.fit(X_train, y_train)

text.insert(END,"Prediction Results\n\n")

 prediction_data = prediction(X_test, cls)

 eml_acc = cal_accuracy(y_test,
prediction_data,'ELM & PSO Algorithm Accuracy,
Classification Report & Confusion Matrix')
```

To evaluate the weight values of genes for samples more effectively, all selected samples in the same class and the different class cover the entire sample dataset as evenly as possible. Since the samples used in the each iteration are all randomly selected, the sample points selected randomly may not be exactly the same as the ReliefF algorithm runs each time, even if the training samples are the same. It follows that the weight values of genes will take on

fluctuation. To solve this issue, the average distance among *k* nearest or *k* non-nearest neighbor samples estimates a quantitative representation of the difference among samples, and many more samples are selected such that it is closer to the actual situation of the samples. It can be observed from Definitions 1 and 2 that the weight fluctuations are efficient, and then the calculation will be more accurate.

Note that when the weight of the important gene becomes larger, it is easily separated from the others and helpful to be selected by the ReliefF algorithm. Meanwhile, when decreasing the distance between the same samples, the distance between the different samples will be increased, so that the difference of weights is very obvious. In order to obtain the more stable results in emergencies, a new distance coefficient is proposed to further reduce the instability during calculations.

The wrapper approach concept involves using learning techniques to select the optimal feature subset [38]. The model hypothesis combines with the classifier in the search space, in order to reach a more accurate classification result. The wrapper technique's quality is gauged by the specific classifier's accuracy. The wrapper approach typically uses evolutionary or bio-inspired algorithms, in order to guide the search process [39]. It first starts with a population of the solution, also known as a feature subset. Next, the features subset is evaluated via the learning strategy, based on the fitness function. Usually, the existence of a different iteration is used, in order to improve the result. The wrapper approach typically requires high computational costs, along with a higher risk of overfitting, while it then shows a better performance than the filter approach [40].

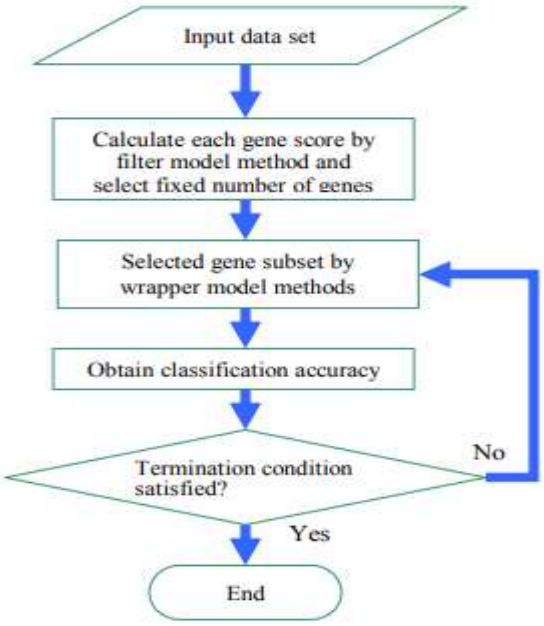


Figure 2. Hybrid filter and wrapper model feature selection method

A correlation based heuristic evaluation function is used to compute the correlation coefficient using the Correlation-based Feature Selection (CFS). It overcomes the disadvantage of univariate filter approaches that does not take into account the interaction between features [15][16]. The identification ability of each of the attributes is used to evaluate a subset of attributes. A multivariate approach is effective in identifying the correlation that exists among the different genes in the dataset [17]. Pearsons correlation coefficient is very sensitive to the presence of outliers and noise[18]. The

relationship between variables (Genes) can be measured by the process of correlation [2]. The linear relationship between two variables in best described in statistics using correlation coefficient or Pearson Product Moment Correlation (PPMC).

Formula for calculating Pearson correlation between features x_i and y_i is given in (1)

$$Correlation = \sum(x_i - mean(x_i)*y_i.mean(y_i) / n*SD(x_i)*SD(y_i)) \quad (1)$$

Pearson correlation coefficient between attributes is found out. Attributes having low inter-correlation are selected [19]. The WEKA tool is used to implement CFS. The selected genes were used to study the different types of cancer. The attributes exhibit high correlation if the value of correlation coefficient lies between 0.5 and 1 and is said to be less correlated if its value lies between 0.3 and 0.5[27].

Particle Swarm Optimization

Particle Swarm Optimization (PSO) is a stochastic evolutionary algorithm based on a population, on the basis of adequate socio-psychological principles to solve engineering problems based on several variables. Swarm Intelligence is embedded in this method. It involves the concept of sharing of information that is simulated by bio inspired behavior. It allows particles to acquire benefits based on discoveries and previous experience, while looking for food. For applications based on PSO, each particle that flies through the search space represents a candidate solution.

The position of a particle is biased by the best position visited using their own knowledge and position of the particle considered by a better knowledge of neighboring particles. When the neighborhood is a swarm of particles, the particle is said to be the world's best particles. The global optimum is measured by a fitness function which varies as a function of the optimization problem [20]. Each particle in the swarm is represented by the following uniqueness:

x_i: current position of the i^{th} particle,

v_i: current velocity of the i^{th} particle,

p_i: best previous position of the i^{th} particle,

gbest: global best particle in its neighborhood.

The personal best position of particle i is the best position experienced by the particle so far. If f is the objective function, the personal best of a particle, at time step Δt is calculated as:

$$p_i(\Delta t+1) = \begin{cases} p_i(\Delta t) & \text{if } f(x_i(\Delta t+1) >= f(p_i(\Delta t)) \\ x_i(\Delta t+1) & \text{if } f(x_i(\Delta t+1) < f(p_i(\Delta t)) \end{cases}$$

If gbest denotes the global best particle, it is given as:

$$gbest(\Delta t)\ C\text{-}\{p_0,p_1,....p_s\} = min\{f(p_0(\Delta t), f(p_1(\Delta t),.... f(p_s(\Delta t))$$

where s is the size of the entire swarm.

The numbers of particles are initialized at random locations that correspond to feature subsets and then swarm towards promising areas via the global best solution so far and each particle's local best. The smallest subset with maximum quality is returned.

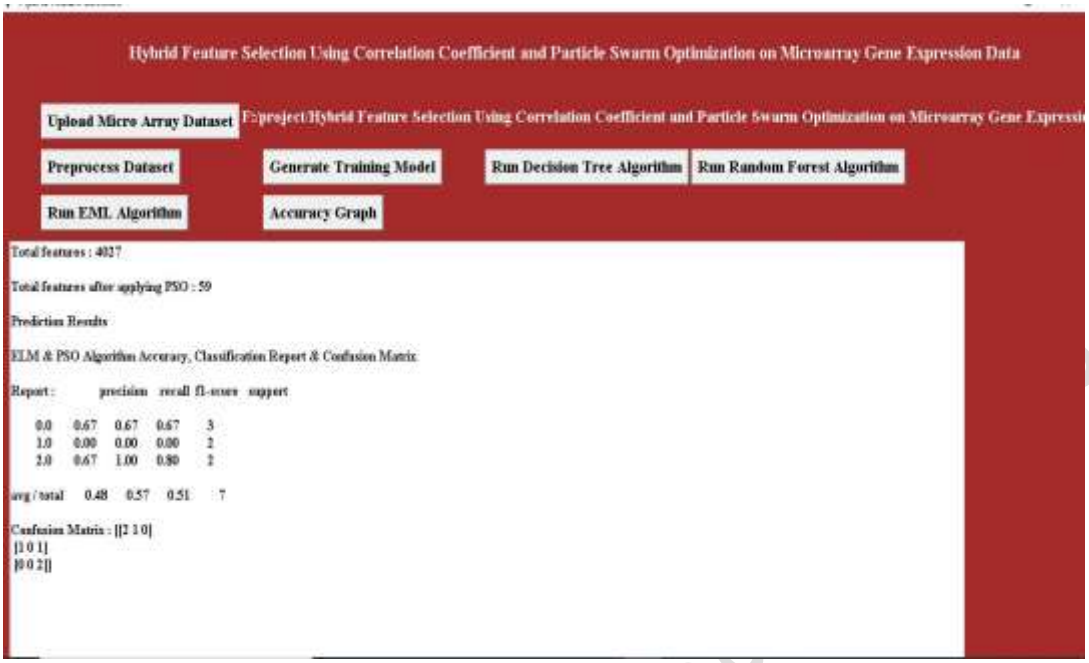
4. RESULTS

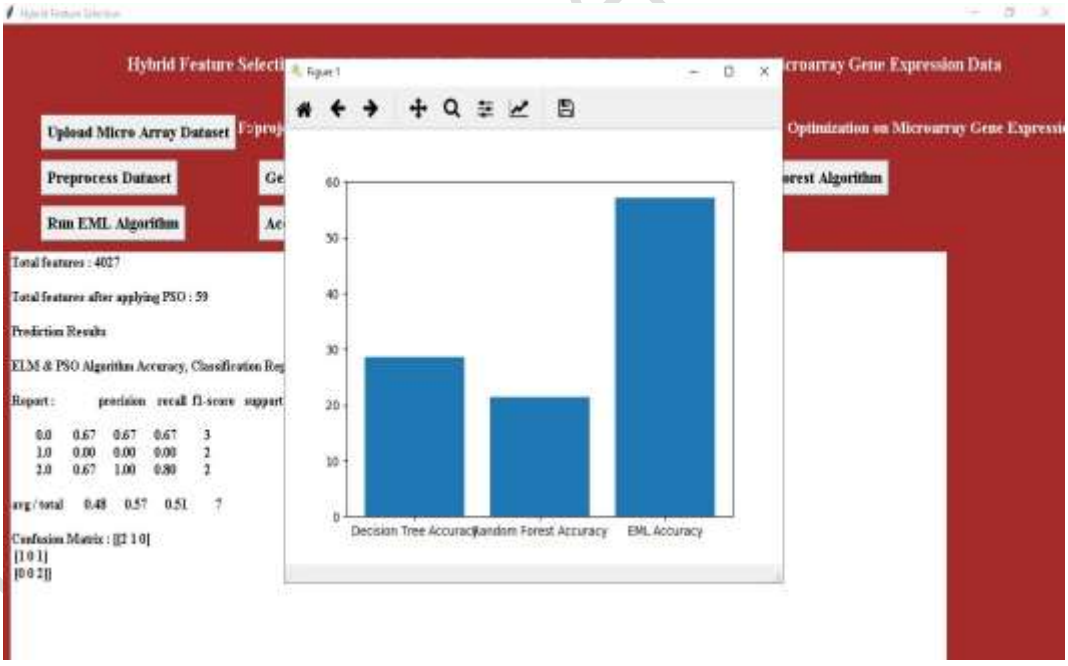
Microarray gene expression data suffers from the problems of missing values due to several experimental reasons. The lymphoma dataset used for our study suffers from this problem. In order to solve this issue, preprocessing is performed on the raw dataset using the impute method. In this case, the missing values are treated using the 'mode' statistical operation wherein the missing values are filled with the value that occurs more often in the dataset. This imputed data is then subjected to feature selection and classification to achieve better classifier accuracy.

Two types of biomedical datasets were used in the present study to investigate how the feature selection methods respond to different data structures. The effectiveness and efficiency of the proposed intelligent hybrid feature selection method was evaluated over six benchmark cancer datasets of microarray gene expression data.

Journal of Engineering Sciences

Vol 11, Issue 4 , April/2020
ISSN NO:0377-9254


Hybrid Feature Selection Using Correlation Coefficient and Particle Swarm Optimization on Microarray Gene Expression Data
Upload Micro Array Dataset
F:/project/Hybrid Feature Selection Using Correlation Coefficient and Particle Swarm Optimization on Microarray Gene Expressi
Preprocess Dataset
Generate Training Model
Run Decision Tree Algorithm
Run Random Forest Algorithm
Run EML Algorithm
Accuracy Graph
Total features : 4027
Total features after applying PSO : 59
Prediction Results
ELM & PSO Algorithm Accuracy, Classification Report & Confusion Matrix
Report : precision recall f1-score support
0.0 0.67 0.67 0.67 3
1.0 0.00 0.00 0.00 2
2.0 0.67 1.00 0.80 2
avg / total 0.48 0.57 0.51 7
Confusion Matrix : [[2 1 0]
[1 0 1]
[0 0 2]]


Upload Micro Array Dataset
Preprocess Dataset
Run EML Algorithm
Total features : 4027
Total features after applying PSO : 59
Prediction Results
Report : precision recall f1-score support
0.0 0.67 0.67 0.67 3
1.0 0.00 0.00 0.00 2
2.0 0.67 1.00 0.80 2
avg / total 0.48 0.57 0.51 7
Confusion Matrix : [[2 1 0]
[1 0 1]
[0 0 2]]
60
50
40
30
20
10
0
Decision Tree Accuracy
Random Forest Accuracy
EML Accuracy

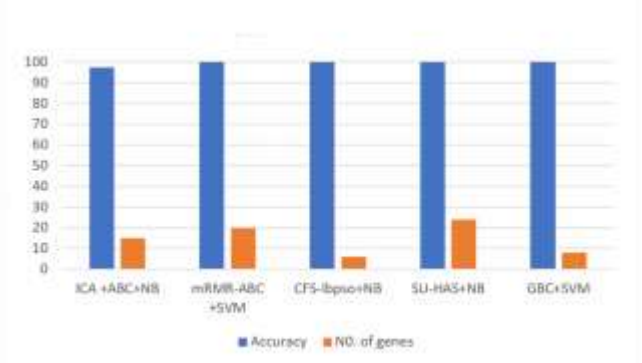


Fig 3: Results of proposed Method

Table 1. human gene expression dataset

Dataset Name	Diagnostic task	Number of		
		Sample	Genes	Classes
9_Tumors	Nine various human Tumor types	60	5726	9
Brain_Tumor1	Five human brain tumor types	90	5920	5
Brain_Tumor2	Four malignant glioma types	50	10367	4
Leukemia1	Acute myelogenous leukemia (AML), acute lymphoblastic leukemia (ALL) B-cell, and ALL T-cell	72	5327	3
Leukemia2	AML, ALL, and mixed-lineage leukemia (MLL)	72	11225	3
DLBCL	Diffuse large B-cell lymphomas and follicular lymphomas	77	5469	2

Table 1. Number of Genes selected using proposed hybrid method

Name of the dataset	**Number of Genes in raw dataset**	**Number of genes selected by proposed method**
SRBCT	2308	75
Lymphoma	4026	286
MLL	12582	1524

5. CONCLUSION

Experimental results showed that the proposed method simplified gene selection and the total number of parameters needed effectively, thereby obtaining a higher classification accuracy compared to other feature selection methods. The classification accuracy obtained by the proposed method was higher than other methods for all six test problems. In the future, the proposed method can assist in further research where feature selection needs to be

implemented. It can potentially be applied to problems in other areas as well. The 10-fold cross validation was used to evaluate the effectiveness of the proposed model and results were compared with five recent recommended hybrid feature selection approaches in literature. Experimental results confirms the superior performance and stability of the proposed hybrid model over other existing methods in terms of the accurate cancer classification and the number of genes selected. Thus the proposed model could be utilized as an efficient tool to reduce gene dimensionality in microarray data analysis. The proposed method significantly reduces the number of genes needed for classification and has also contributed to the improvement in classifier accuracy. The proposed method has greater scope of application to problems in other domains in future and this research will be extended to investigate the applicability of proposed model with other high dimensional datasets and improve the efficiency in terms of computational complexity. In this work, the classifier accuracy of the proposed hybrid approach that combines the correlation coefficient with particle swarm optimization.. It is evident that the extreme learning machines classifier produces more or comparatively better accuracy than the other tree based classifiers available in literature. The proposed hybrid method that has higher potential in aiding further research in the area of feature selection simplified the process of gene selection which is evident from the experimental results. The proposed method significantly reduces the number of genes needed for classification and has also contributed to the improvement in classifier accuracy. The proposed method has greater scope of application to problems in other domains in future.

REFERENCES

[1]. Lakshman Narayana Vejendla and A Peda Gopi, (2019),” Avoiding Interoperability and Delay in Healthcare Monitoring System Using Block Chain Technology”, Revue d'Intelligence Artificielle , Vol. 33, No. 1, 2019,pp.45-48.

[2]. Gopi, A.P., Jyothi, R.N.S., Narayana, V.L. et al. (2020), “Classification of tweets data based on polarity using improved RBF kernel of SVM” . Int. j. inf. tecnol. (2020). https://doi.org/10.1007/s41870-019-00409-4.

[3]. A Peda Gopi and Lakshman Narayana Vejendla, (2019),” Certified Node Frequency in Social Network Using Parallel Diffusion Methods”, Ingénierie des Systèmes d' Information, Vol. 24, No. 1, 2019,pp.113-117.. DOI: 10.18280/isi.240117

[4]. Lakshman Narayana Vejendla and Bharathi C R ,(2018),“Multi-mode Routing Algorithm with Cryptographic Techniques and Reduction of Packet Drop using 2ACK scheme in MANETs”, Smart Intelligent Computing and Applications, Vo1.1, pp.649-658. DOI: 10.1007/978-981-13-1921-1_63 DOI: 10.1007/978-981-13-1921-1_63

[5]. Lakshman Narayana Vejendla and Bharathi C R, (2018), “Effective multi-mode routing mechanism with master-slave technique and reduction of packet droppings using 2-ACK scheme in MANETS”, Modelling, Measurement and Control A, Vol.91, Issue.2, pp.73-76. DOI: 10.18280/mmc_a.910207

[6]. Lakshman Narayana Vejendla , A Peda Gopi and N.Ashok Kumar,(2018),“ Different techniques for hiding the text information using text steganography techniques: A survey”, Ingénierie des Systèmes d'Information, Vol.23, Issue.6,pp.115-125.DOI: 10.3166/ISI.23.6.115-125

[7]. A Peda Gopi and Lakshman Narayana Vejendla (2018), “Dynamic load balancing for client server assignment in distributed system using genetic algorithm”, Ingénierie des Systèmes d'Information, Vol.23, Issue.6, pp. 87-98. DOI: 10.3166/ISI.23.6.87-98

[8]. Lakshman Narayana Vejendla and Bharathi C R,(2017),“Using customized Active Resource Routing and Tenable Association using Licentious Method Algorithm for secured mobile ad hoc network Management”, Advances in Modeling and Analysis B, Vol.60, Issue.1, pp.270-282. DOI: 10.18280/ama_b.600117

[9]. Lakshman Narayana Vejendla and Bharathi C R,(2017),“Identity Based Cryptography for Mobile ad hoc Networks”, Journal of Theoretical and Applied Information

Technology, Vol.95, Issue.5, pp.1173-1181. EID: 2-s2.0-85015373447

[10]. Lakshman Narayana Vejendla and A Peda Gopi, (2017)," Visual cryptography for gray scale images with enhanced security mechanisms", Traitement du Signal,Vol.35, No.3-4,pp.197-208. DOI: 10.3166/ts.34.197-208

[11]. A Peda Gopi and Lakshman Narayana Vejendla, (2017)," Protected strength approach for image steganography", Traitement du Signal, Vol.35, No.3-4,pp.175-181. DOI: 10.3166/TS.34.175-181

[12]. Lakshman Narayana Vejendla and A Peda Gopi, (2020)," Design and Analysis of CMOS LNA with Extended Bandwidth For RF Applications", Journal of Xi'an University of Architecture & Technology, Vol. 12, Issue. 3,pp.3759-3765. https://doi.org/10.37896/JXAT12.03/319.

[13]. Chaitanya, K., and S. Venkateswarlu,(2016),"DETECTION OF BLACKHOLE & GREYHOLE ATTACKS IN MANETs BASED ON ACKNOWLEDGEMENT BASED APPROACH." Journal of Theoretical and Applied Information Technology 89.1: 228.

[14]. Patibandla R.S.M.L., Kurra S.S., Mundukur N.B. (2012), "A Study on Scalability of Services and Privacy Issues in Cloud Computing". In: Ramanujam R., Ramaswamy S. (eds) Distributed Computing and Internet Technology. ICDCIT 2012. Lecture Notes in Computer Science, vol 7154. Springer, Berlin, Heidelberg

[15]. Patibandla R.S.M.L., Veeranjaneyulu N. (2018), "Survey on Clustering Algorithms for Unstructured Data". In: Bhateja V., Coello Coello C., Satapathy S., Pattnaik P. (eds) Intelligent Engineering Informatics. Advances in Intelligent Systems and Computing, vol 695. Springer, Singapore

[16]. Patibandla, R.S.M.L., Veeranjaneyulu, N. (2018), "Performance Analysis of Partition and Evolutionary Clustering Methods on Various Cluster Validation Criteria", Arab J Sci Eng ,Vol.43, pp.4379–4390.

[17]. R S M Lakshmi Patibandla, Santhi Sri Kurra and N.Veeranjaneyulu, (2015), "A Study on Real-Time Business Intelligence and Big Data",Information Engineering, Vol.4,pp.1-6.

[18]. K. Santhisri and P.R.S.M. Lakshmi,(2015), " Comparative Study on Various Security Algorithms in Cloud Computing", Recent Trends in Programming Languages ,Vol.2,No.1,pp.1-6.

[19]. K.Santhi Sri and PRSM Lakshmi,(2017), "DDoS Attacks, Detection Parameters and Mitigation in Cloud Environment", IJMTST,Vol.3,No.1,pp.79-82.

[20]. P.R.S.M.Lakshmi,K.Santhi Sri and Dr.N. Veeranjaneyulu,(2017), "A Study on Deployment of Web Applications Require Strong Consistency using Multiple Clouds", IJMTST,Vol.3,No.1,pp.14-17.

[21]. P.R.S.M.Lakshmi,K.Santhi Sri and M.V.Bhujanga Ra0,(2017), "Workload Management through Load Balancing Algorithm in Scalable Cloud", IJASTEMS,Vol.3,No.1,pp.239-242.

[22]. K.Santhi Sri, P.R.S.M.Lakshmi, and M.V.Bhujanga Ra0,(2017), "A Study of Security and Privacy Attacks in Cloud Computing Environment", IJASTEMS,Vol.3,No.1,pp. 235-238.

[23]. R S M Lakshmi Patibandla and N. Veeranjaneyulu, (2018), "Explanatory & Complex Analysis of Structured Data to Enrich Data in Analytical Appliance", International Journal for Modern Trends in Science and Technology, Vol. 04, Special Issue 01, pp. 147-151.

[24]. R S M Lakshmi Patibandla, Santhi Sri Kurra, Ande Prasad and N.Veeranjaneyulu, (2015), "Unstructured Data: Qualitative Analysis", J. of Computation In Biosciences And Engineering, Vol. 2,No.3,pp.1-4.

[25]. R S M Lakshmi Patibandla, Santhi Sri Kurra and H.-J. Kim,(2014), "Electronic resource management using cloud computing for libraries", International Journal of Applied Engineering Research, Vol.9,pp. 18141-18147.

[26]. Ms.R.S.M.Lakshmi Patibandla Dr.Ande Prasad and Mr.Y.R.P.Shankar,(2013), "SECURE ZONE IN CLOUD", International Journal of Advances in Computer Networks and its Security, Vol.3,No.2,pp.153-157.

[27]. Patibandla, R. S. M. Lakshmi et al., (2016), "Significance of Embedded Systems to IoT.", International Journal of Computer Science and Business Informatics, Vol.16,No.2,pp.15-23.

[28]. AnveshiniDumala and S. PallamSetty. (2020),"LANMAR routing protocol to support real-time communications in MANETs using Soft computing technique", 3rd International Conference on Data Engineering and Communication Technology (ICDECT-2019), Springer, Vol. 1079, pp. 231-243.

[29]. AnveshiniDumala and S. PallamSetty. (2019),"Investigating the Impact of Network Size on LANMAR Routing Protocol in a Multi-Hop Ad hoc Network", i-manager's Journal on Wireless Communication Networks (JWCN), Volume 7, No. 4, pp.19-26.

[30]. AnveshiniDumala and S. PallamSetty. (2019),"Performance analysis of LANMAR routing protocol in SANET and MANET", International Journal of Computer Science and Engineering (IJCSE) – Vol. 7,No. 5, pp.1237-1242.

[31]. AnveshiniDumala and S. PallamSetty. (2018), "A Comparative Study of Various Mobility Speeds of Nodes on the Performance of LANMAR in Mobile Ad hoc Network", International Journal of Computer Science and Engineering (IJCSE) – Vol. 6, No. 9, pp. 192-198.

[32]. AnveshiniDumala and S. PallamSetty. (2018),"Investigating the Impact of IEEE 802.11 Power Saving Mode on the Performance of LANMAR Routing Protocol in MANETs", International Journal of Scientific Research in Computer Science and Management Studies (IJSRCSMS) – Vol.7, No. 4.

[33]. AnveshiniDumala and S. PallamSetty. (2016),"Analyzing the steady state behavior of RIP and OSPF routing protocols in the context of link failure and link recovery in Wide Area Network", International Journal of Computer Science Organization Trends (IJCOT) – Vol. 34 No 2, pp.19-22.

[34]. AnveshiniDumala and S. PallamSetty. (2016),"Investigating the Impact of Simulation Time on Convergence Activity & Duration of EIGRP, OSPF Routing Protocols under Link Failure and Link Recovery in WAN Using OPNET Modeler", International Journal of Computer Science Trends and Technology (IJCST) – Vol. 4 No. 5, pp. 38-42.

[35]. VellalacheruvuPavani and I. Ramesh Babu (2019) ,"Three Level Cloud Storage Scheme for Providing Privacy Preserving using Edge Computing",International Journal of Advanced Science and Technology Vol. 28, No. 16, pp. 1929 – 1940.

[36]. VellalacheruvuPavani and I. Ramesh Babu,"A Novel Method to Optimize the Computation Overhead in Cloud Computing by Using Linear Programming" ,*International Journal of Research and Analytical Reviews* May 2019, Volume 6, Issue 2,PP.820-830..

[37]. Anusha Papasani and Nagaraju Devarakonda,(2016),"Improvement of Aomdv Routing Protocol in Manet and Performance Analysis of Security Attacks", International Journal Of Research in Computer Science & Engineering ,Vol.6,No.5, pp.4674-4685.

[38]. Sk.Reshmi Khadherbhi,K.Suresh Babu , Big Data Search Space Reduction Based On User Perspective Using Map Reduce ,International Journal of Advanced Technology and Innovative Research Volume.07, IssueNo.18, December-2015, Pages: 3642-3647

[39]. B.V.Suresh kumar,Sk.Reshmi Khadherbhi ,BIG-IOT Framework Applications and Challenges: A Survey Volume 7, Issue VII, JULY/2018 pg.no 1257-1264

[40]. P.Sandhya Krishna,Sk.Reshmi Khadherbhi,V.Pavani, Unsupervised or Supervised Feature Finding For Study of Products Sentiment ,International Journal of Advanced Science and Technology, Vol 28 No 16 (2019).

[41]. K.Santhi Sri, Dr.Ande Prasad (2013), "A Review of Cloud Computing and Security Issues at Different Levels in Cloud Computing" , International Journal on Advanced Computer Theory and Engineering Vol. 2,pp 67-73.

[42]. K.Santhi Sri, N.Veeranjaneyulu(2018), "A Novel Key Management Using Elliptic and Diffie-Hellman for Managing users in Cloud Environment", Advances in Modelling and Analysis B,Vol.61,No.2,pp 106-112.

[43]. K.Santhi Sri, N.Veeranjaneyulu(2019), "Decentralized Key Management Using Alternating Multilinear Forms for Cloud Data Sharing with Dynamic Multiprivileged Groups", Mathematical Modelling of Engineering Problems,Vol.6,No.4,pp511-518.

[44]. S.Sasikala, P.Sudhakar, "interpolation of CFA color Images with Hybrid image denoising", 2014 Sixth International Conference on Computational Intelligence and Communication Networks, DOI 10.1109/.53 193 DOI 10.1109/CICN.2014.53, pp. 193-197.

[45]. Me. Jakeera Begum and M.Venkata Rao, (2015), "Collaborative Tagging Using CAPTCHA" International Journal of Innovative Technology And Research, Volume No.3, Issue No.5,pp,2436 – 2439.

[46]. L.Jagajeevan Rao, M. Venkata Rao, T.Vijaya Saradhi (2016), "How The Smartcard Makes the Certification Verification Easy" Journal of Theoretical and Applied Information Technology, Vol.83. No.2, pp. 180-186.

[47]. Venkata Rao Maddumala, R. Arunkumar, and S. Arivalagan (2018)"An Empirical Review on Data Feature Selection and Big Data Clustering" Asian Journal of Computer Science and Technology Vol.7 No.S1, pp. 96-100.

[48]. Singamaneni Kranthi Kumar, Pallela Dileep Kumar Reddy, Gajula Ramesh, Venkata Rao Maddumala, (2019), "Image Transformation Technique Using Steganography Methods Using LWT Technique" ,Traitement du Signalvol 36, No 3, pp. 233-237.