Journal of Theoretical and Applied Information Technology
15th November 2021. Vol.99. No 21

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

A HYBRID FEATURE SELECTION BASED ADVANCED NEURAL NETWORK OPTIMIZATION MODEL ON SOMATIC CANCER PREDICTION

P. R.SUDHA RANI[1], Dr. K. KIRAN KUMAR[2]

[1]Research Scholar, Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, AP, India. sudharani.p@gmail.com

[2]Professor, Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, AP, India. kiran5434@kluniversity.in

ABSTRACT

As the size of the micro-array databases and its dimensions are increasing, the prediction of different gene based disease rules are difficult to handle due to large features space in the biomedical applications. Micro-array data play important roles in the different disease pathogenesis. However, experimental prediction of microRNAs interactions with disease is still difficult. In addition, some essential limitations of previous computer methods are identified in the disease network to classify possible interactions of gene-disease. Most of the conventional feature selection approaches and classification models are used to predict the gene based patterns on limited dimension space and data size. In this work, a hybrid feature selection measure is developed to find the essential key features for the classification problem. In this work, a filter based feature selection and hybrid classification approaches are implemented on the high dimensional features. Experimental study show that the present model has better gene-based pattern extraction efficiency and runtime(ms) than the conventional models.

Keywords: *Gene Based Micro-Array Dataset, Feature Selection Measures, Gene Classifiction Disease Prediction.*

1. INTRODUCTION

Classification is the supervised learning technique in which the data sample is classified into an existing classification. A classification device finds the rule that assigns an existing class or label to a new sample[1]. The concept of classification of data applies to many patterns such as high dimensional data analysis, natural language processing, handwriting recognition, computer biology, text mining, drug design, patient identification of diseases and many more. The classification algorithm creates a classifier in the training phase which analyses the training samples with the characteristics and the associated labels. This is called supervised learning, because class labels are available beforehand. In the second stage, the performance of the classifier is assessed using the test samples. The test samples are broken down from the training samples and randomly selected from the whole dataset. Various performance measures are used to evaluate the test result[2].

Supervised techniques of selecting features, however, involve adequate marked information that involve comprehensive knowledge and are costly to acquire. Semi-supervised choice techniques use label data and data allocation information or local composition of both marked and unlabeled data to assess function relevance[3]. There is no extensive study on semi-supervised techniques of selecting features, however.

Microarray data set is one of the most critical chronic diseases, and evolving from mild symptoms to severe disease and death takes a long period[4]. Medical data is generally containing a set of cancer patches and its disease associations. It is very much difficult for the doctor to detect at high risk of having this disease[5]. Apart from this, patients' knowledge and previous clinical history assists them to detect the disease easily. Many researchers have been using statistical and mining tools to help doctors detect cancer disease[6]. In the analysis of cancer disease, satisfactory levels of results with better accuracy can only be achieved with mining

algorithms. Many of the researchers examined hybrid methods to obtain improved results in the prediction of cancer diseases, involving many algorithms. No more consideration was given to finding the correct treatment procedure for patients with cancer disease. To overcome the disadvantages, the gap between identifying the disease and the necessary treatments has been identified [7-9]. This method analyses whether true performance is supported by the treatment dataset as well as finding the cancer-related disease. Classification techniques are implemented annually by researchers for the diagnosis of cancer-related diseases due to the increase in death rates of diseased people and the availability of large quantities of patient information[10].

Variable stratification is used to help train classifiers with uneven data distributions. It uses two different classification methods. The evolutionary algorithm assigns base probabilities to each classifier in a dynamic manner. The second method employs the use of a neural classifier to fuse the results of the base classifiers. There was considerable agreement between the methods used, and the overall effectiveness of the predictions has increased since. This is attempting to employ a classification strategy that minimizes error. Malignant cases were found to be easily and rapidly identified using this method. In other words, this system performs a complex analysis of the SVM and its various kernel functions. This SVM ensemble was utilized as the base classifiers, and the separate kernel functions for each classifier. The base classifiers are learned using either a small or large database, depending on their size. The individual classifiers' predictions are combined and the final judgment was reached by averaging those results. In order to measure the performance of the individual SVM model and the method accuracy, ROC, Fmeasure, and computational time were utilized Two techniques were employed in this plan, a bagging approach and boosting. Other testing has shown that SVM ensembles produce better results the model self-governing feature selection algorithms can differentiate between various classifiers

With regard to the dataset, the result totally depends on the class label. In this case, they use the fisher feature-centered evaluation as the starting point for feature ranking problem. Alternative criteria may be used to evaluate the relevancy of a feature. Choosing features based on the eigenvectors or Laplian score . Then in the article [11], they uncovered many of the features overlooked by prior research was brought to light; after that study, a large number features ranking measures were identified for classification problem.

For the estimator, a standard norm is added to the cost. As a result, a sparser vector will result in most features taking on the lower side of the range. [12] proposed a graph lasso, as a possible alternative. It is described as being resourceful and scalable because it is separate from classifiers. They lend themselves to this model. However, if the classifier is known, then the classifier wrapper model may not be applicable. Some well-known filter algorithms include the following: Information Gain, t-test, FCBF, Chi-Square Correlation Matrix, and Gini Index. Classifier ensembles use recursive partitioning algorithms, which divide the space into regions that include only one class of instances. The project deals with simple, multivariate perception tree algorithms that make a decision about each node on the scale of 1 to 7 and are able to handle not only yes/no queries. If the subspace can be linearly separated by linear methods, it tries a heuristic measure first. In case of separable subsets, a linear dimension is used.

In this paper, the main contributions include:

1.Proposed a hybrid feature selection measure for the gene feature extraction and ranking.

2. Implemented a hybrid gene based disease classification model to improve the error rate and runtime(ms).

2. RELATED WORK

A DNA microarray analysis generates thousands of measurements of gene expression and an easy way to collect enormous amount of data in a short period of time[13]. Profiles of gene expression are more objective, accurate and reliable than traditional diagnostic approaches[14]. But when dealing with these data for data mining, different problems arise. One of the biggest drawbacks is the size of these databases. In addition, there is not enough sample to train and validate the developed models. Another weakness of microarray data is that of the large number of genes, few are significant and that the rest is redundant, noisy or much less important. The presence of these genes degrades the classification and decreases prediction performance while increasing the error rate[15]. Most importantly, we need human specialists to identify a collection of information. Unlabeled

information, on the other side, are often abundant and comparatively simple to acquire and cheap to get.

A meta-heuristic algorithm performs optimization and always delivers near optimal results, if not perfectly results in less number of iterations. The meta-analysis results show promise, whether or not preprocessing has been done on the raw data. PSO has made significant advances into global meta-heuristic approaches to solving meta-heuristic problems[16]. It is similar to various species' behavior, like ant, birds, fish, in that it uses pheromones. When particles or members of the swarm carry out their multidimensional search, they search the space of solutions instead of just in one dimension at a time. Each particle recalibrates its position from time to the total available search space with flying experience, according to each to the other particles in the field. The PSO algorithm is always improving the candidates as they are being discovered. Many possible solutions exist to achieve an optimal solution, the particles move at a specified velocity,each particle maintains a memory trace of its previous location, so it can return to its preferred location[17]. The particles in the communication process determine the local best and global best. Over time, PSO has found successful application in many different scientific studies. The major benefit of this work is the use of a PSO algorithm for classification problems on massive datasets. It is tried in the current experiment to explore the possibility of hybrid optimization. It does amazing things by itself, but it also lends itself to a more refined computational analysis, which is beneficial when used in conjunction with other techniques[18]. The use of parallel kernel computing techniques improves computational performance, for instance, by taking advantage of the parallel simulation of multiple Marcov chains on every node. AdaBoost, Negative-Correlation learning, and genetic ensemble gave better generalization abilities than the other approaches. They used meta-heuristic techniques to find optimal solutions such as Ant Colony Optimization (ACO), Chromosome reduction, and genetic algorithms can also be used to tackle metaheurism problems. Convergent thinking may bring you closer to the global optimum, but there are no guarantees of convergence. Another new approach to metaheuristic algorithms like the Dragonfly algorithm, binary gray optimization, and the whale algorithm also outweighted existing techniques by balancing between the exploration and exploitation phases[19]. Additionally, several authors have used hybrid meta-heuristic techniques to improve the mixing of features of two or more meta-heuristic approaches in their work. Exploration is critical to ensure search over every part of the solution space in order to provide an accurate global optimum, and exploitation is key to where we find better solutions by applying local search[20].

The main gaps identified in this paper include

1. Most of the models are difficult to improve the error rate due to high dimensionality.
2. Problem to handle high dimensional features using correlation measures.
3. Difficult to improve the overall true positive rate.

The main findings in this paper include:

1. Implementing a hybrid correlated feature selection measure for the classification problem.
2. Developing a hybrid classification approach using the weighted neural network.

Pros and Cons of conventional models:

Reference	Advantages	Problem/ Research gap
[16]	Easy to handle limited feature space	Difficult to improve the error rate due to high dimensionality
[17]	Better runtime compared to the traditional models on limited datasets	Difficult to improve the error rate due to high dimensionality
[18]	Better applicable to small datasets with limited features.	Problem to handle high dimensional features using correlation measures
[19]	Easy to handle limited feature space	Difficult to improve the overall true positive rate.
[20]	Better runtime compared to the traditional models on limited datasets	Difficult to improve the overall true positive rate.

3. PROPOSED MODEL

In this work, a hybrid framework is developed to improve the overall gene based disease prediction on large micro-array databases. The overall framework of the proposed model is summarized in fig 1.

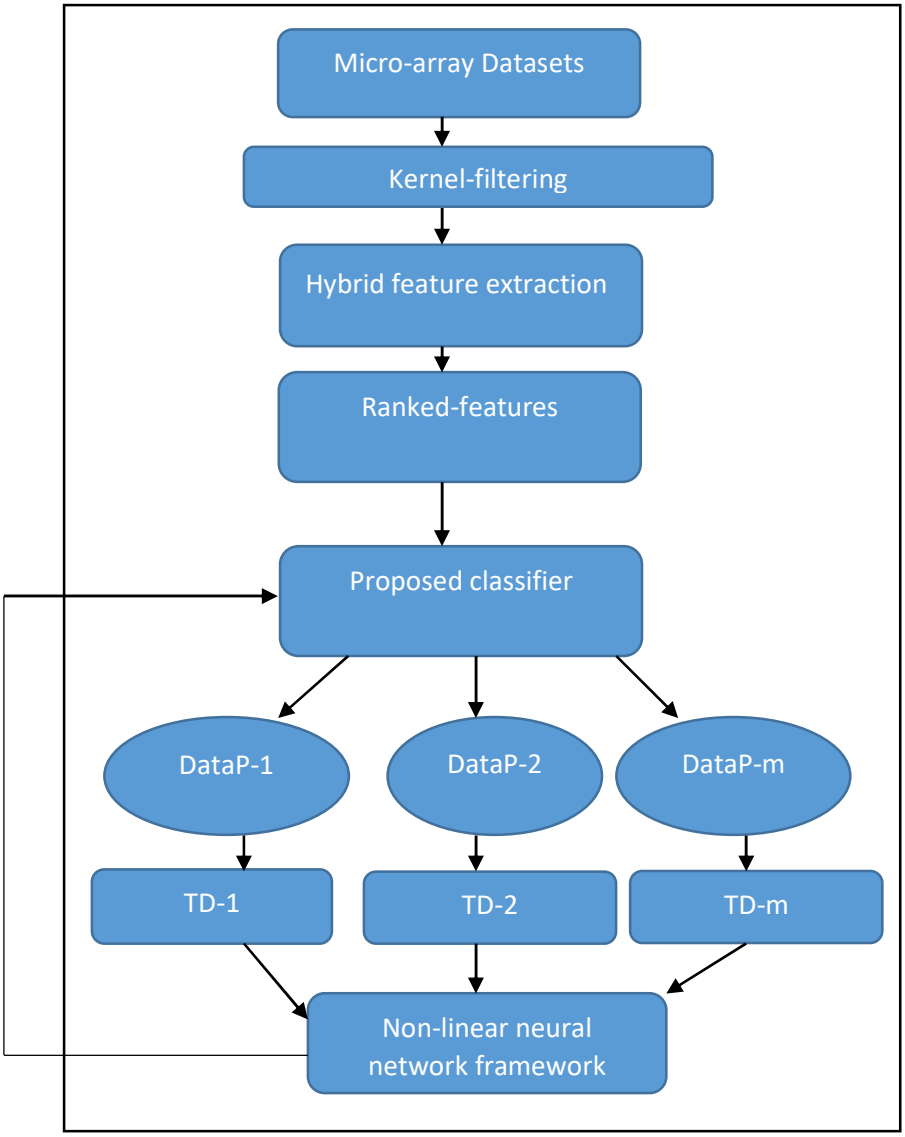


Figure 1: Proposed Kernel Filter Based Non-Linear Neural Network

Microarray training data is pre-processed using data transformation functions to reduce noise in data distributions. Clustering and wrapper feature ranking are performed using a kernel-based data transformation function in the mapper phase of the proposed work.

Journal of Theoretical and Applied Information Technology
15th November 2021. Vol.99. No 21

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Kernel based Data filtering

Input : Gene based disease dataset D.

Output: Filtered dataset.

Steps:

1. Get micro-array dataset D.
2. To each pair of dimensions in D
3. do
4. Perform data filtering using kernel approach as
5. $$Gene\_Kernel\_Transform(Fl[i], F2[j]) = \frac{1}{1+\eta^2 / \cos(\max(\sigma^2_{F[i]}, \sigma^2_{F[j]})}$$
6. Where $$\eta = 2 * \sum Feat[i].Feat[j] - \sum (Feat[i] + Feat[j])^2$$
7. If($Gene\_Kernel\_Transform(Fl[i], F2[j]) >$ Thres)
8. Then
9. Normalize Feat[i] and Feat[j] within $[0, GeneKernelTransform(Feat[i], Feat[j])]$ using Min-max normalization as KD
10. Else
11. Apply min-max normalization
12. End if
13.
14. Compute Similarity between gene symbol and gene name using the similarity measure.
15. $$Gsim[] = \max\{\frac{1}{3}\{\alpha(\frac{1}{|FS(D,i)|} + \frac{1}{|GD[j]|}) + (1-\frac{t}{\beta})\}, \frac{|FS(D,i) \cap GD[j]|}{|FS(D,i) \cap GD[j]| + \alpha|FS(D,i) - GD[j]| + \beta|GD[j] - FS(D,i)|}\}$$
16. Done

Here , minimum and maximum scaling is used to normalize the values in the specified range.

Boosting feature selection based hybrid classifier

Most of the conventional models are difficult to filter the essential key features due to high computational memory . In this work, a hybrid weighted feature selection based classification model is proposed for efficient classification process.

1: Load dataset D

2: Read number of clusters c.

4: Initialize random centroids as k.

5: To each cluster in K

Compute the clustering between the features using correlational measure. Let feature vector one V1, feature vector 2 V2

$$Dist(V1, V2) = \frac{\sqrt{Cosine(V1[i], V2[i])}}{Correlation(V1, V2).\sqrt[3]{\sum V1[i]^2 + \sum V2[i]^2}}$$

Done

6: Update cluster centroid using mean distance.

Proposed classification model

Step 1: Assign feature weighted using the maximized weights using the (1) ,(2) and (3) to each feature. Using the standard deviation of the class labels, the T-statistical weighting measure is used to find the variation in the gene characteristics. It is basically the ratio of the class label to the maximized standard deviation.

$$W1 = \frac{\mu_P - \mu_N}{\sqrt{\max\{\sigma^2_{c1}/|P|, \sigma^2_{c2}/|N|\}}} \quad \text{-----(1)}$$

where μ_{c1} is the mean of the class instances.

$$W2 = \frac{|\mu_{ci} - \mu_{cj}|}{2(\sigma_{c1} + \sigma_{c2})} \quad \text{-----(2)}$$

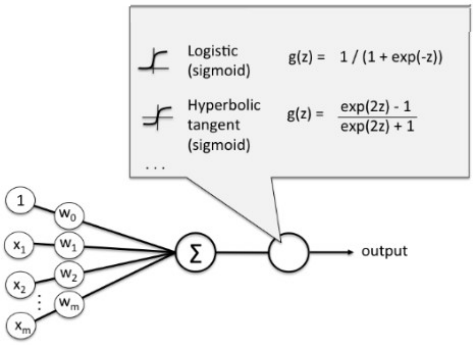
where μ_P and σ_P are the averge and standard-deviation class samples.

$$W3 = \max\{Correlation(ClusterFeatures : CF), \frac{\mu_{c1} - \mu_{c2}}{\sqrt{\max\{\sigma^2_{c1}/|P|, \sigma^2_{c2}/|N|\}}}, \frac{|\mu_{c1} - \mu_2|}{2(\sigma_{c1} + \sigma_{c2})}\} \quad \text{-----(3)}$$

Weights W[]=Max{W1,W2,W3}

Step 2: Here, the number of input , hidden and output layers are defined prior to error computation in each step.

Step 3: In the hidden layer, the weights and activation function are applied to optimize the overall network framework.



Step 4: Repeat this process till error rate and weights are optimized for true positive prediction.

4. EXPERIMENTAL RESULTS

To evaluate the model's performance in existing models, different microarrays from the biomedical repository have been selected. Table 1 summarizes different data sets used for the experimental evaluation.

In this study, proposed feature ranking based classification approach improves the overall feature selection and error rate during the model initialization and classification process. In this work, the high dimensional feature space is taken as input to find the correlated features and classification problem. From the experimental analysis, it is clear that the present model has better error rate than the conventional models for biomedical databases.

Table. 1 Datasets And Its Characteristics

Micro array Datasets	Gene sets	Data-Type
Prostate	2136	Continuous/Numeric
Lymphoma	5000	Continuous/Numeric
DLBCL-Stanford	4000	Continuous/Numeric
Breast cancer	24481	Continuous/Numeric
Leukemia	7129	Continuous/Numeric

In this work, proposed model optimizes the overall true positive rate and error rate than the state of art algorithms. Since, proposed model is implemented on high dimensional feature space, the overall efficiency of the model increases with size.

Cancer Data Results

A. Ovarian Cancer

```
MZ14855.232 < 0.19
|  MZ17187.862 < 0.14
|  |  MZ6610.4417 <0.45 : Cancer (6/0)
|  |  MZ6610.4417 >= 0.45 : Normal (1/0)
|  MZ17187.862 >= 0.14 : Normal (36/0)
MZ14855.232 >= 0.19
|  MZ10624.907 < 0.31
|  |  MZ8060.174 < 0.56
|  |  |  MZ37.485529 < 0.56
|  |  |  |  MZ14271.872 < 0.26
|  |  |  |  |  MZ13649.995 <0.17 : Cancer (31/0)
|  |  |  |  |  MZ13649.995 >= 0.17
|  |  |  |  |  |  MZ1585.0452 <0.35 : Normal (4/0)
|  |  |  |  |  |  MZ1585.0452 >= 0.35
|  |  |  |  |  |  |  MZ0.21817145 <0.47 : Normal (2/0)
|  |  |  |  |  |  |  MZ0.21817145 >= 0.47 : Cancer (6/0)
|  |  |  |  MZ14271.872 >= 0.26 : Cancer (89/0)
|  |  |  MZ37.485529 >= 0.56 : Normal (5/0)
|  |  MZ8060.174 >= 0.56
|  |  |  MZ30.219974 <0.27 : Cancer (2/0)
|  |  |  MZ30.219974 >= 0.27
|  |  |  |  MZ275.34532 <0.69 : Normal (9/0)
|  |  |  |  MZ275.34532 >= 0.69 : Cancer (1/0)
|  MZ10624.907 >= 0.31
|  |  MZ257.67477 < 0.41
|  |  |  MZ4064.1028 < 0.33
|  |  |  |  MZ13240.959 <0.74 : Cancer (5/0)
|  |  |  |  MZ13240.959 >= 0.74
|  |  |  |  |  MZ11.733841 <0.52 : Normal (2/0)
|  |  |  |  |  MZ11.733841 >= 0.52 : Cancer (1/0)
|  |  |  MZ4064.1028 >= 0.33
|  |  |  |  MZ17.797652 <0.58 : Normal (30/0)
|  |  |  |  MZ17.797652 >= 0.58
|  |  |  |  |  MZ177.1166 <0.61 : Cancer (2/0)
|  |  |  |  |  MZ177.1166 >= 0.61 : Normal (1/0)
|  |  MZ257.67477 >= 0.41
|  |  |  MZ52.994969 <0.18 : Normal (1/0)
|  |  |  MZ52.994969 >= 0.18 : Cancer (19/0)
```

B. MLL Cancer

```
2057_g_at < 300.5
|  40869_at < 737
|  |  1184_at <7167 : AML (14/0)
|  |  1184_at >= 7167
|  |  |  34204_at <21.5 : MLL (2/0)
|  |  |  34204_at >= 21.5 : AML (2/0)
|  40869_at >= 737
|  |  35296_at <813 : MLL (12/0)
|  |  35296_at >= 813
|  |  |  39615_at < 917
|  |  |  |  34132_at < -74 : ALL (4/0)
|  |  |  |  34132_at >= -74 : MLL (1/0)
|  |  |  39615_at >= 917
|  |  |  |  35543_at < -132.5 : AML (1/0)
|  |  |  |  35543_at >= -132.5 : MLL (2/0)
2057_g_at >= 300.5
|  41431_at <92.5 : AML (3/0)
|  41431_at >= 92.5 : ALL (16/0)
```

Lung-cancer Michigan

Attribute	Statistical Results
A28102_at	
mean	148.2313
std. dev.	119.2281
AB000114_at	
mean	84.0833
std. dev.	52.8439
AB000115_at	
mean	326.5354
std. dev.	231.216
AB000220_at	
mean	569.7792
std. dev.	479.704
AB000381_s_at	
mean	19.7917
std. dev.	35.6906
AB000409_at	
mean	341.6667
std. dev.	131.6691
AB000410_s_at	
mean	-18.3896
std. dev.	103.0914
AB000449_at	
mean	111.0917
std. dev.	95.6018
AB000450_at	
mean	273
std. dev.	148.9532
AB000460_at	
mean	888.9708
std. dev.	463.2539
AB000462_at	
mean	137.0271
std. dev.	87.0154
AB000464_at	
mean	213.6708
std. dev.	96.635
AB000466_at	
mean	-78.1438
std. dev.	78.1981
AB000467_at	
mean	18.3896
std. dev.	60.5334
AB000468_at	
mean	745.9812
std. dev.	261.0011
AB000584_at	
mean	1351.4542
std. dev.	1417.5612
AC002077_at	
mean	343.9458
std. dev.	185.1255
AC002086_at	
mean	43.9417
std. dev.	41.6335
AC002115_cds1_at	
mean	4658.4312
std. dev.	1202.2574
AC002115_cds3_at	
mean	407.4437
std. dev.	309.3359
AC002115_cds4_at	
mean	214.0438
std. dev.	135.8106
AC002115_rna2_at	
mean	429.325
std. dev.	164.6693

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

AC002450_at	
mean	99.7708
std. dev.	68.2991
AC002464_at	
mean	12.0167
std. dev.	33.6334
AC002477_s_at	
mean	281.0188
std. dev.	84.5061
AC002486_at	
mean	76.6354
std. dev.	43.8582
AD000092_cds1_at	
mean	132.2062
std. dev.	62.0227
AD000092_cds2_at	
mean	370.1396
std. dev.	101.1167
AD000092_cds7_s_at	
mean	305.6833
std. dev.	104.3436
AD000684_cds1_at	
mean	671.4667
std. dev.	448.4904
AD001527_cds1_at	
mean	133.7458
std. dev.	106.5897
AF000177_at	
mean	403.8188
std. dev.	183.685
AF000231_at	
mean	180.1396
std. dev.	107.415

AF000234_at	
mean	243.1792
std. dev.	117.1593
AF000424_s_at	
mean	251.2625
std. dev.	339.5603
AF000430_at	
mean	55.1104
std. dev.	51.8018
AF000545_at	
mean	200.2417
std. dev.	82.0901
AF000560_at	
mean	230.2875
std. dev.	94.0669
D13969_at	
mean	32.4125
std. dev.	109.3532
D13988_at	
mean	909.9146
std. dev.	307.6384
D14043_at	
mean	988.4917
std. dev.	408.206
D14134_at	
mean	4.9583
std. dev.	26.3696
D14446_at	
mean	589.1542
std. dev.	1314.0003
D14497_at	
mean	34.8229
std. dev.	37.7081
D14520_at	

mean	447.9062
std. dev.	328.0253
D14530_at	
mean	12365.2458
std. dev.	2682.5123
D14533_at	
mean	107.3437
std. dev.	41.4744
D14657_at	
mean	412.9083
std. dev.	363.6342
D14658_at	
mean	2474.0708
std. dev.	667.2929

Table 3:Performance Analysis Of Proposed Model To The Traditional Models On Three Microarray Datasets By Using Accuracy Measure

Datasets	PSO + CART	PSO + SVM	PSO + FFNN	IPSO + Ensemble	Proposed classifier
Lung Cancer	0.714	0.8353	0.85	0.9674	0.975
Lung Michigan	0.814	0.8353	0.84	0.914	0.99
Ovarian	0.835	0.8744	0.85	0.964	0.984
Prostate Tumor	0.846	0.8835	0.89	0.981	0.989
Error Rate	0.314	0.3042	0.29	0.17	0.191
Runtime (ms)	6983	6194	6364	4893	4785

The suggested model's performance on all cancer datasets is shown in Table 3. Cancer datasets are analysed in this study to establish an average true positive and precision rate for the high dimensional datasets. There can be clearly shown that the current approach outperforms the current models in terms of true positive rate and precision.

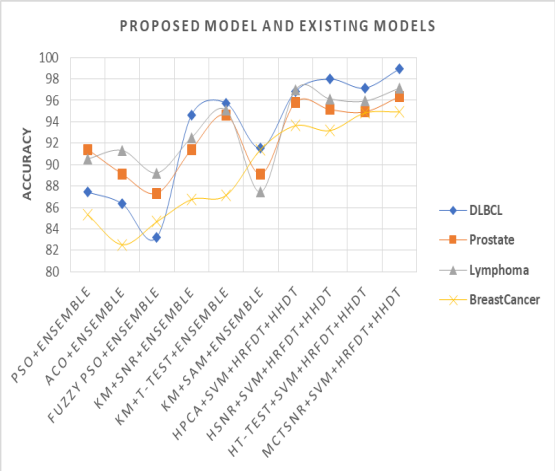


Figure 2:Comparison Of Proposed Model To Existing Models On Average Feature Seleciton.

Interpretation:

From the experimental results and analysis , it is observed that the present model has better accuracy on the high dimesnional datasets with nearly 2% improvement over the conventioanl models. Also, the runtime of the model is optimized nearly 5% over the conventioanl models.

Limitations of the work:

The main limitations of the work include:

1. Difficult to classify large datasize with morethan 10 lakhs instances due to high computational memory.
2. Difficult to handle multi-class classification problem.

5. CONCLUSION

In this paper, a hybrid feature selection based classifier is implemented on the high dimensional biomedical datasets. Since, most of the traditional approaches are difficult to predict the disease based on the correlational features due to high computational time and conventional ranking measures. . In this work, a hybrid feature selection measure is developed to find the essential key features for the classification problem. In this work, a filter based feature selection and hybrid classification approaches are implemented on the high dimensional features. Experimental study show that the present model has better gene-based pattern extraction efficiency and runtime(ms) than the conventional models. In future work, a parallel cluster based classification model is proposed on the feature space to improve the heterogeneous disease types.

References

[1] S. Belciug and F. Gorunescu, "Learning a single-hidden layer feedforward neural network using a rank correlation-based strategy with application to high dimensional gene expression and proteomic spectra datasets in cancer detection", Journal of Biomedical Informatics 83 (2018) 159–166.

[2] V. Bolón-Canedo, N. Sánchez-Maroño, A. Alonso-Betanzos, J. M. Benítez and F. Herrera , A review of microarray datasets and applied feature selection methods, Information Sciences 282 (2014) 111–135.

[3] V. Bolón-Canedo, N. Sánchez-Maro˜no and A. Alonso-Betanzos, Distributed feature selection: An application to microarray dataclassification, Applied Soft Computing 30 (2015) 136–150.

[4] S. H. Bouazza, K. Auhmani, A. Zeroual and N. Hamdi, Selecting significant marker genes from microarray data by filter approach for cancer diagnosis, Procedia Computer Science 127 (2018) 300–309.

[5] S. Chormungea and S. Jena, Correlation based feature selection with clustering for high dimensional data, Journal of Electrical Systems and Information Technology, 2017.

[6] R. Dash, A Two Stage Grading Approach for Feature Selection and Classification of Microarray Data using Pareto based Feature Ranking Techniques: A Case Study, Journal of King Saud University - Computer and Information Sciences.

[7] M. Ghosh , S. Begum , R. Sarkar , D. Chakraborty and U. Maulik, Recursive Memetic Algorithm for Gene Selection in Microarray Data, Expert Systems with Applications.

[8] S. Guo, D. Guo, L. Chen and Q. Jiang, A L1-regularized feature selection method for local dimension reduction on microarray data, Computational Biology and Chemistry 67 (2017) 92–101.

[9] Q. Hou, Z. Bing, C. Hu, M. Li , K. Yang , Z. Mo, X. Xie, J. Liao, Y. Lu, S. Horie and M. Lou, RankProd Combined with Genetic Algorithm Optimized Artificial Neural Network Establishes a Diagnostic and Prognostic Prediction Model that Revealed C1QTNF3 as a Biomarker for Prostate Cancer

[10] Y. Liu , Prominent feature selection of microarray data, Progress in Natural Science 19 (2009) 1365–1371. [16] Bangare S.L., Pradeepini G., Patil S.T. (2017),'Brain tumor classification using mixed method approach',2017 International Conference on Information Communication and Embedded Systems, ICICES 2017.

[11] Potharaju S.P., Sreedevi M. (2017),'A novel M-cluster of feature selection approach based on symmetrical uncertainty for increasing classification accuracy of medical datasets',Journal of Engineering Science and Technology Review,10(6),PP.154-162.

[12] Angel Prathyusha K., Mahitha Y., Prasanna Kumar Reddy N., Raja Rajeswari P. (2018),'A survey on prediction of suitable crop selection for agriculture development using data mining classification techniques',International Journal of Engineering and Technology(UAE),7 (0),PP. 107-109

[13] Potharaju S.P., Sreedevi M. (2018),'A Novel Subset Feature Selection Framework for Increasing the Classification Performance of SONAR Targets',Procedia Computer Science,125 (),PP. 902-909

[14] Sajana T., Narasingarao M.R. (2018),'Classification of imbalanced malaria disease using naïve bayesian algorithm',International Journal of Engineering and Technology(UAE),7 (),PP. 786-790

[15] Tayar Y., Ram Prasad R.S., Satayanarayana S. (2018),'An accurate classification of imbalanced streaming data using deep convolutional neural network',International Journal of Mechanical Engineering and Technology,9 (3),PP. 770-783

[16] Kuchibhotla S., Niranjan M.S.R. (2018),'Emotional classification of Acoustic information with optimal feature subset selection methods',International Journal of Engineering and Technology(UAE),7 (2),PP. 39-43

[17] Masoodi, Tariq Ahmad; Shaik, Noor Ahmad; Burhan, Syed; Hasan, Qurratulain; Shafi, Gowhar; Talluri, Venkateswara Rao(2019),` Structural prediction, whole exome sequencing and molecular dynamics simulation confirms p.G118D somatic mutation of PIK3CA as functionally important in breast cancer patients', COMPUTATIONAL BIOLOGY AND CHEMISTRY,80,PP.472-479.DOI: 10.1016/j.compbiolchem.2019.05.012

[18] Thimmaraju, Manish Kumar; Bheemanapally, Khaggeswar; Dharavath, Rajkumar; Kakarla, Lavanya; Botlagunta, Mahendran(2017),` Improved Anticancer Activity of Meloxicam Hydrogels in K562 and HL60 Cell Lines', JOURNAL OF YOUNG PHARMACISTS,9(2),PP.209-213.DOI: 10.5530/jyp.2017.9.41

[19] Sai, KKS; Bashetti, N; Chen, XF; Norman, S; Hines, JW; Meka, O; Kumar, JVS; Devanathan, S; Deep, G; Furdui, CM; Mintz, A(2019),`Initial biological evaluations of F-18-KS1, a novel ascorbate derivative to image oxidative stress in cancer', EJNMMI RESEARCH, DOI:10.1186/s13550-019-0513-x.

[20] Sivakumar, S; Nayak, SR; Vidyanandini, S; Kumar, JA; Palai, G(2018),` An empirical study of supervised learning methods for breast cancer diseases',OPTIK, DOI:10.1016/j.ijleo.2018.08.112