An Efficient Approach to Microarray Data Classification using Elastic Net Feature Selection, SVM and RF

C Koushik¹, Shreyas Madhav A V² and Rabindra Kumar Singh¹
¹,²,³ Vellore Institute of Technology (VIT) Chennai, Tamil Nadu, India
E-mail: rabindrakumar.singh@vit.ac.in

Abstract. DNA Microarray technology forms an integral part of the bioinformatics world and provides researchers with the ability to monitor a large amount of gene expressions simultaneously. The analysis of this data proves extremely beneficial in the detection of several diseases. Classification of the gene expression data obtained from microarrays is an imperative step in providing further information on the types of possible diseases present. Several classification models for microarray data have been proposed, yielding considerable results in terms of accuracy and execution time. This paper demonstrates the implementation of a microarray data classification system using Elastic Net for feature selection on two classification mechanisms. The classification of the data has been carried out using SVM and Random Forest Classifier methods. The implemented system using Elastic Net and SVM exhibits a much better performance in terms of accuracy and execution time than most of the existing systems.

1. Introduction

Microarray data has been regarded as one of the most popular ways for various disease identifications, primarily cancer in the field of biological computations [1]. Biologists have been provided with the ability to calculate and monitor the expression levels of millions of genes within a single observation due to microarray data. Several statistical challenges arise when analyzing the large amount of gene expression data obtained, namely in classifying the datasets elements into their respective classes. Predicting the classes for new samples based on identification of the gene data is our main goal. In classification of microarray gene data, the main motive is to derive a prominent and robust relationship between the given set of genes, which come under the same class label and use this information to predict and classify genes encountered in the future. In a more mathematical sense, this problem can be stated as development of a classifier that can provide a mapping between the object and its relevant class label. A classifier is better defined as a classification algorithm that enforces a concrete implementation. Considering the huge amount of gene expression data involved for processing in retrieving disease information, classification of the micro dataset plays a major step in determining the overall working rate and functioning of the disease detection process. Hence, a well-structured classification system that provides accurate results in minimal time is necessary in computing the biological information present. Several machine learning systems are currently available and multiple approaches that have been employed to solve this problem. The classification problem consists of two major steps that determine the viability of the outcome: Feature selection mechanism and a classifier mechanism.

Feature Selection is an imperative process of any classification problem where features that most efficiently help categorize and label the input data by contributing to the prediction output that is required for the user. The feature selection process is responsible for the identification of differentially expressed genes that have the potential to help segregate relevant data into their respective categories. A very high gene pool that is present initially should be reduced significantly. A gene pool of 8000-
11000 should be brought down to 200-300. The elimination of unnecessary features that do not contribute to the classification process is as imperative as selecting the most relevant features. We can achieve this by explicit feature selection or by feature weighting. If the feature selection process is not done properly, irrelevant, and unnecessary features may be taken in account for training of the model and will result in very poor performance of the classification system. At this point, it is essential that we understand the difference between feature extraction and feature selection. Feature extraction aims to create new features by deriving functions of the originally available features in the data, whereas feature selection focusses on retrieving a subset of the original features. Feature selection techniques are most often used in applications where the amount of features is high and the number of samples or data points available is comparatively lesser. Several feature selection processes have been established in the past namely, Elastic Net [2], LASSO [3] and Ridge [4] and each feature selection methodology holds its own advantages and disadvantages based upon the application.

Classification plays a vital role in the field of Machine Learning and Statistics. It is a supervised learning method that classify new observations based on previous learning done by the system which was executed on input learning data provided by the programmer. One prominent algorithm that has been used in several biological computational problems is the Support Vector Machine. This algorithm has been proven effective for both classification and regression problems. Plotting data items as points in a space with dimensions equal to the number of features we have, classification in SVM is based upon locating the hyper plane that best separates the two classes. This hyper-plane supervised learning method provides apt solutions in problems of classification and regression. The SVM model focusses on maximizing the margin and works towards minimizing the structural risks that come into play. The model evaluates small and large datasets with very little performance variation and hence, we consider it to the perfect candidate for this application. For the sake of comparing the SVM model, we implement the random forest classification algorithm for the same algorithm and draw a parallel between their performances. The random forest classification algorithm is a standard method that is employed in various current day scenarios. The random forest classifier fits multiple decision trees for classification on subsets of the dataset and performs averaging as a method for reducing over fitting of the model and improve accuracy of the system.

2. Material and Methods
The aim of this paper is to predict gene expression classes across multiple microarray datasets using classification. The same is achieved by training machine learning based classification prediction models on the existing microarray datasets and perform comparative studies between the models based on their overall accuracy and execution speed. This section discusses the datasets used and methods proposed for classification of DNA genes using large arrays of microarray expressions.

2.1. Datasets
The datasets used in this paper containing DNA gene data including diseases like Crohn’s Disease, Glioma, Huntington’s Disease, various forms of cancer and healthy gene data (subject to aging and natural effects). The dataset files are in csv file formats and accessed by our program initially for the feature selection process. These are well known datasets that have been used for classification studies in several previous papers and provide an all-round presence of various disease data.

We consider a total of 16 previously established datasets that provide high dimensional and small sample type of DNA microarray data, namely Chin[5], Chowdry[6], Gravier[7], Sorlie[8], West[9], Pomeroy[10], Burczynski[11], Sun[12], Borovecki[13], Chiaretti[14], Singh[15], Nakayama[16], Khan[17], Subramanian[18], Gordon[19] and Tian[20]. Table 1 provides a detailed description
of the contents of each dataset. The datasets have been subjected to feature selection process to identify the most relevant features that contribute towards the classification process.

**Table 1. Datasets and their properties**

| Dataset       | Disease            | No. of Samples | No. of Features | Classes |
|---------------|--------------------|----------------|-----------------|---------|
| Chin (2006)   | Breast Cancer      | 118            | 22215           | 2       |
| Chowdary (2006)| Breast Cancer      | 104            | 22283           | 2       |
| Gravier (2010)| Breast Cancer      | 168            | 2905            | 2       |
| Sorlie (2001) | Breast Cancer      | 85             | 456             | 5       |
| West (2001)   | Breast Cancer      | 49             | 7129            | 2       |
| Pomeroy (2002)| CNSET              | 60             | 7128            | 2       |
| Burcyznski (2006)| Crohn’s Disease   | 127            | 22283           | 3       |
| Sun (2006)    | Glioma             | 180            | 54613           | 4       |
| Borovecki (2005)| Huntington’s Disease | 31          | 222839          | 2       |
| Chiaretti (2004)| Leukaemia         | 128            | 12625           | 2       |
| Singh (2002)  | Prostate Cancer    | 102            | 12600           | 2       |
| Nakayama (2007)| Sarcoma           | 105            | 22283           | 10      |
| Khan (2001)   | SRBCT              | 63             | 2308            | 4       |
| Subramanian (2005) | N/A              | 50             | 10100           | 2       |
| Gordon (2002) | Lung Cancer        | 181            | 12533           | 2       |

Each dataset is of different type and properties, resulting in a wide array of features selected. The accuracy and the execution time required is also calculated on each dataset’s classification process.

**3. Proposed Method**

The proposed approach to the problem has two main components to it: Feature Selection and Classification. The brief overview of the steps required to create this system are as follows. The Elastic Net feature selection algorithm is applied to all the datasets and the feature count and total features that are selected from each dataset are stored. We now make use of our SVM model for classification of the results.

A Random forest based classification is also executed from the feature count obtained from Elastic Net as a point of comparison for our SVM model. Figure 1 represents the overall idea of this proposed system framework.
3.1. Feature Selection

In microarray data classification, selecting the most appropriate feature selection method for a given problem is required for the system to identify the most relevant features and produce accurate results. The feature selection process starts off by extraction of the features and their labels from the csv files. The value places containing N/A are replaced with zero. The next step involves conversion of the features and labels into NumPy arrays, followed by the process of feature scaling. An appropriate scaling approach is employed here. The next step, which is the most crucial step of the process, is selection/elimination of features using Elastic Net. We have selected as the learning-rate parameter for Elastic Net as 0.005. Once the elastic fit has been carried out, it is necessary for us to save the feature scores for usage in further processing. The scores are stored to a text document and accessed during the SVM stage. The total selected feature count is analysed, and it has been found that our proposed system leads to a much higher number of selected features than previous systems. This feature selection plays a critical role in distinguishing our system and providing the increase in accuracy.

3.2. Elastic Net

We use Elastic Net for feature selection for both the classification methods that have been compared in this paper-Support Vector Machine and Random forest Classification. Elastic Net is a well-known regularised linear regression method that works based on the combination of the L1 and L2 penalties. It can be described as an extension of the original linear regression method which adds penalties of regularization during the process of training. It works particularly well when we encounter the presence of lots of unnecessary variables that need to be taken out of the equation and works well when we have a lot of useful variables that are essential for the process and need to be retained. Other
situations where this method proves beneficial is when we need to handle correlated variables in the process. The Elastic Net follows a shrinkage and selection approach that proves more efficient in performance when compared to the LASSO feature selection model that is used most frequently in existing applications. This approach provides us with an increase in predictive power compared to the LASSO feature and can viewed as providing the combined benefits of both LASSO and Ridge methods. In order to provide a more detailed description, we can say that Elastic Net combines the feature group selection characteristics of ridge with the feature selection properties of LASSO. This varied feature selection method contributes to much higher feature selection possibilities leading to higher accuracy in a much lesser time frame.

3.3. LOOCV Cross Validation

Next, we perform train-test split on the data. Sometimes, the split may not be random and may lead to overfitting where the predicted line captures almost every single point that cannot explain the variability, or underfitting when the line does not capture enough points to predict an outcome, leading to huge bias and error values. Therefore, it is important to cross validate the split to evaluate performance of models on unknown data. For this reason, we use the Leave One Out Cross Validator (LOOCV) to split the data into multiple folds that are averaged and fit in the model. Now that our dataset is comparatively smaller, having large number of folds will help reduce the error due to bias.

LOOCV is also the most popular procedure for estimating the performance of the model based upon the predictions made by the system on data that have not been used in the training process. Every evaluation made is finally summed up and averaged to attain the final score. We have implemented this cross validation methodology in our proposed system for the performance evaluation and validation analysis.

3.4. Classification

Now we pass on the results to the SVM classifier. A linear kernel with penalty of L2 was employed for the SVM model. The SVM model deployed follows similar principles and parameters as used in existing systems. The final accuracy and time required for total execution for each dataset. An iteration count maxed at 1000 and the loss function was calculated using the squared hinge loss. This accuracy and time parameters obtained are checked against similar proposed systems of the past to demonstrate the improved efficiency of our system.

We also implement a random forest approach to the same problem to draw a comparison between the SVM and the original RF implementation used. An estimator count of 10 for selection and 2 as the maximum count. We do not achieve a increase in accuracy with the RF model but the execution time is affected by the parameters. When we compare our RF model to previous RF models that implement LASSO as their feature selection process, we see a clear increase in the accuracy and time taken by our system.

In this paper, we establish a model with a unique combinational system of Elastic Net and SVM, which results in higher accuracy in much lesser time when compared to the previously implemented models. We apply the feature selection process of Elastic Net to the Random Forest classifier as well for comparing it with previous random forest systems and prove the efficiency of Elastic Net in the system.

4. Experimental Results

We establish a comparison of our proposed approach using Elastic Net and SVM primarily with the original model presented by Güçkıran K et al[21]. and also with other similar models that have been established in the past. There have been various studies conducted with several approaches on this classification problem but most of them revolve around only one or two datasets. We have used studies
that provide a more generalized approach and cover multiple datasets for comparison. The statistics and results mentioned for every study is their best outcomes. Our proposed approach surpasses all the compared approaches by providing equal or higher accuracy in similar or lesser execution time. The implementation of the feature selection and classification process was carried out using Python using library frameworks such as Keras, TensorFlow, SciPy, NumPy, Pandas and Matplotlib. Table 2 describes the accuracy and execution times required to classify various well-known datasets. A comparison with the original SVM model based on LASSO feature selection is done to show the efficiency of the Elastic Net feature selection process. The following experiments were carried out on an i7-8550U CPU @ 1.80GHz (8 CPUs) with 16GB of RAM.

Table 2. Accuracy and Execution Time comparison of SVM models

| Dataset  | SVM LASSO Accuracy | SVM LASSO Selected Features | SVM ElasticNet Execution Time (secs) | SVM ElasticNet Accuracy | SVM ElasticNet Selected Features | SVM ElasticNet Execution Time (secs) |
|----------|--------------------|-----------------------------|--------------------------------------|-------------------------|---------------------------------|--------------------------------------|
| Chin     | 100.00%            | 112                         | 3.4908                               | 100.00%                 | 132                             | 3.2126                               |
| Chowdary | 100.00%            | 83                          | 3.6081                               | 100.00%                 | 106                             | 2.4696                               |
| Gravier  | 98.20%             | 128                         | 4.1562                               | 99.40%                  | 175                             | 6.0944                               |
| Sorlie   | 84.80%             | 74                          | 1.5477                               | 89.40%                  | 90                              | 1.4867                               |
| West     | 94.20%             | 51                          | 0.1971                               | 100.00%                 | 64                              | 0.1944                               |
| Pomeroy  | 100.00%            | 56                          | 0.2936                               | 100.00%                 | 68                              | 0.3223                               |
| Burczynski | 93.50%             | 122                         | 5.0530                               | 95.27%                  | 157                             | 4.4405                               |
| Sun      | 83.30%             | 181                         | 36.3164                              | 83.30%                  | 226                             | 32.4293                              |
| Borovecki| 100.00%            | 30                          | 0.1706                               | 100.00%                 | 50                              | 0.1600                               |
| Chiaretti| 96.10%             | 123                         | 5.1405                               | 96.10%                  | 154                             | 4.8692                               |
| Singh    | 100.00%            | 84                          | 1.7004                               | 100.00%                 | 104                             | 1.6655                               |
| Nakayama | 69.70%             | 129                         | 5.0948                               | 73.30%                  | 218                             | 4.4557                               |
| Khan     | 100.00%            | 56                          | 0.3679                               | 100.00%                 | 72                              | 0.4031                               |
| Subramanian | 100.00%         | 46                          | 0.4224                               | 100.00%                 | 59                              | 0.2865                               |
| Gordon   | 100.00%            | 97                          | 3.5878                               | 100.00%                 | 134                             | 3.7371                               |

We see that the proposed model performs very well and provides an average accuracy of 95.8 for all the datasets in a total time of 66.2269 secs which is a significant improvement when compared to the original accuracy of 94.65% in a total time of 71.1473. Table 3 carries out a similar comparison between the RF models that have been implemented in the previous system and in our proposed system. We see that an overall average accuracy of 86.6% has been achieved across all datasets with execution time of 84.7961 seconds which is a great improvement when compared to the original RF LASSO system, which presents a average accuracy of 81.9% with a total time of 167.4897. We observe that the execution time is halved in both classification scenarios while maintaining equal or achieving much higher accuracy.
Table 3. Accuracy and Execution Time comparison of RF models

| Dataset   | RF LASSO Accuracy | Selected Features | RF LASSO Execution Time (secs) | RF ElasticNet Accuracy | Selected Features | RF ElasticNet Execution Time (secs) |
|-----------|-------------------|-------------------|--------------------------------|------------------------|-------------------|-------------------------------------|
| Chin      | 90.7%             | 112               | 11.5910                         | 90.7%                  | 132               | 5.6861                              |
| Chowdary  | 98.1%             | 83                | 9.1072                          | 98.1%                  | 106               | 4.1240                              |
| Gravier   | 74.4%             | 128               | 13.0430                         | 82.7%                  | 175               | 4.8270                              |
| Sorlie    | 62.4%             | 74                | 5.5455                          | 81.2%                  | 90                | 1.5464                              |
| Pomeroy   | 66.7%             | 56                | 4.0201                          | 75.0%                  | 68                | 1.1560                              |
| Burczynski| 75.6%             | 122               | 12.3288                         | 79.5%                  | 157               | 6.4247                              |
| Sun       | 60.0%             | 181               | 40.9372                         | 70.6%                  | 226               | 30.5523                             |
| Borovecki | 100%              | 30                | 1.9995                          | 100%                   | 50                | 0.6560                              |
| Chiaretti | 81.3%             | 123               | 12.0164                         | 87.5%                  | 154               | 5.8111                              |
| Singh     | 94.1%             | 84                | 7.6545                          | 95.1%                  | 104               | 2.8274                              |
| Nakayama  | 58.1%             | 129               | 9.8571                          | 63.8%                  | 218               | 4.9388                              |
| Khan      | 100%              | 56                | 4.0147                          | 100%                   | 72                | 1.0935                              |
| Subramanian| 88.0%             | 46                | 3.2179                          | 94.0%                  | 59                | 0.9841                              |
| Gordon    | 99.4%             | 97                | 16.0079                         | 99.4%                  | 134               | 6.5474                              |
| Tian      | 79.2%             | 158               | 16.1489                         | 81.5%                  | 181               | 7.7013                              |

Table 4 provides an overall analysis of our system with several previously established systems, that have been tested and implemented with a couple of datasets for specific applications. Our model which provides a generalized classification approach proves much more efficient and viable, surpassing them in terms of both accuracy and execution time. This comparative analysis helps study our proposed system and its capabilities.

Table 4. Results

| Dataset   | Ari. [22] | Hue. [23] | Phu. [24] | Mun. [25] | Le. [26] | Güç [20] | Proposed System Accuracy |
|-----------|-----------|-----------|-----------|-----------|---------|---------|--------------------------|
| Chin      | 88.0%     | N/A       | N/A       | N/A       | N/A     | 100%    | 100%                     |
| Chowdary  | 95.20%    | 97.30%    | 98.10%    | N/A       | N/A     | 100%    | 100%                     |
| Gravier   | 76.20%    | 75.00%    | 79.80%    | N/A       | N/A     | 98.20%  | 99.4%                     |
| Sorlie    | N/A       | N/A       | N/A       | N/A       | N/A     | 84.70%  | 89.4%                     |
| West      | N/A       | N/A       | N/A       | N/A       | N/A     | 100%    | 100%                     |
| Pomeroy   | 63.30%    | N/A       | N/A       | N/A       | N/A     | 92.10%  | 100%                     |
| Burczynski| 86.50%    | N/A       | N/A       | N/A       | N/A     | 93.50%  | 95.3%                     |
| Sun       | N/A       | N/A       | N/A       | N/A       | N/A     | 72.30%  | 83.3%                     |
Figure 2 describes comparison of the accuracy achieved by the SVMs for each dataset. Figure 3 provides a similar comparison on the two systems based on their RF model’s performance. Figure 4 and Figure 5 plot the average accuracy and overall execution times to validate our proposed approach.

|         | SVM LASSO Accuracy | SVM ElasticNet Accuracy | RF LASSO Accuracy | RF ElasticNet Accuracy |
|---------|--------------------|-------------------------|-------------------|-----------------------|
| Borovecki | N/A                | N/A                     | 99.20%            | 99.70%                |
| Chiaretti | N/A                | N/A                     | 96.30%            | 96.50%                |
| Singh    | 99.50%             | 97.20%                  | N/A               | 100%                  |
| Nakayama  | N/A                | N/A                     | 89.90%            | 73.3%                 |
| Khan      | N/A                | N/A                     | 99.60%            | 100%                  |
| Subramanian | N/A              | N/A                     | N/A               | 100%                  |
| Gordon    | 99.40%             | N/A                     | 100%              | 100%                  |

Figure 2. SVM Accuracy comparison

Figure 3. Random Forest Accuracy comparison
Through graphical tools, we aim to provide a vivid understanding and comparative analysis of our robust model’s performance with the existing systems. From figures 4 and 5, we notice that, on average, combining all the datasets, our system performs significantly better in terms of both average accuracy and overall execution time.

5. Conclusion

The proposed implementation in this paper provided a general approach to microarray data classification that has been tested using a variety of different datasets. The system considers and extracts higher number of features, resulting in higher accuracy and less execution time compared to most of the previous approaches. The SVM classifier in this implementation exhibits better performance in terms of both accuracy and execution time when compared to the Random Forest Classifier. Although usage of SVMs for classification of DNA microarray data is widely used across the field of computational biology, the Elastic Net feature selection based method proposed in this paper proves to be a more viable solution for producing accurate results and completing the tasks in time less than or equal to the time required in the previously proposed systems. Further work can be done in improving the system by parallelization of the SVM model, providing data augmentation.
features to the model or expanding the model to much larger datasets for further improvement and validation.

References

[1] Sanchez A. and de Villa M.C.R. 2008 A tutorial review of microarray data analysis (Department d’Estadstica. Universitat de Barcelona)

[2] Zou, Hui & Hastie, Trevor. (2005). Regularization and variable selection via the Elastic Net (vol B 67, pg 301, 2005). Journal of the Royal Statistical Society Series B. 67. 768-768. 10.1111/j.1467-9868.2005.00527.x.

[3] Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society. Series B (Methodological), 267-288.

[4] Arthur E. Hoerl & Robert W. Kennard (1970) Ridge Regression: Biased Estimation for Non-Orthogonal Problems, Technometrics, 12:1, DOI: 10.1080/00401706.1970.10488634

[5] Chin, K., DeVries, S., Fridlyand, J., Spellman, P. T., Roydasgupta, R., Kuo, W. L., ... & Chen, F. (2006). Genomic and transcriptional aberrations linked to breast cancer pathophysiology. Cancer cell, 10(6), 529-541.

[6] Chowdary, D., Lathrop, J., Skelton, J., Curtin, K., Briggs, T., Zhang, Y., & Mazumder. A. (2006) Prognostic gene expression signatures can be measured in tissues collected in RNAlater preservative. The Journal of Molecular Diagnostics, 8(1), 31-39.

[7] Gravier, E., Pierron, G., Vincent Salomon, A., Guel, N., Raynal, V., Savignoni, A., ... & Forquet A. (2010). A prognostic DNA signature for T1T2 nodenegative breast cancer patients. Genes, chromosomes and cancer, 49(12), 1125-1134.

[8] Sørlie, T., Perou, C. M., Tibshirani, R., Aas, T., Geisler, S., Johnsen, H., ... & Thorsen, T. (2001) Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proceedings of the National Academy of Sciences, 98(19), 10869-10874.

[9] West, M., Blanchette, C., Dressman, H., Huang, E., Ishida, S., Spang, R., ... & Nevins, J. R. (2001). Predicting the clinical status of human breast cancer by using gene expression profiles. Proceedings of the National Academy of Sciences, 98(20), 11462-11467.

[10] Pomeroy, S. L., Tamayo, P., Gaasenbeek, M., Sturla, L. M., Angelo, M., McLaughlin, M. E., & Allen, J. C. (2002). Prediction of central nervous system embryonal tumour outcome based on gene expression. Nature, 415(6870), 436.

[11] Burczynski, M. E., Peterson, R. L., Twine, N. C., Zuberek, K. A., Brodeur, B. J., Casciotti, L., ... & Spinelli, W. (2006). Molecular classification of Crohn’s disease and ulcerative colitis patients using transcriptional profiles in peripheral blood mononuclear cells. The Journal of Molecular Diagnostics, 8(1), 51-61.

[12] Sun, L., Hui, A. M., Su, Q., Vortmeyer, A., Khotiarov, Y., Pastorino, S., ... & Rosenblum, M. (2006). Neuronal and glioma-derived stem cell factor induces angiogenesis within the brain. Cancer cell, 9(4), 287-300.

[13] Borovecki, F., Lovrecic, L., Zhou, J., Jeong, H., Then, F., Rosas, H. D., ... & Krainc, D. (2005). Genome-wide expression profiling of human blood reveals biomarkers for Huntington’s disease. Proceedings of the National Academy of Sciences, 102(31), 11023-11028.

[14] Chiaretti, S., Li, X., Gentleman, R., Vitale, A., Vignetti, M., Mandelli, F., ... & Foa, R. (2004). Gene expression profile of adult T-cell acute lymphocytic leukemia identifies distinct subsets of patients with different response to therapy and survival. Blood, 103(7), 2771-2778.
[15] Singh, D., Febbo, P. G., Ross, K., Jackson, D. G., Manola, J., Ladd, C., ... & Lander, E. S. (2002). Gene expression correlates of clinical prostate cancer behavior. Cancer cell, 1(2), 203-209.

[16] Nakayama, R., Nemoto, T., Takahashi, H., Ohta, T., Kawai, A., Seki, K., ... & Hasegawa, T. (2007). Gene expression analysis of soft tissue sarcomas: characterization and reclassification of malignant fibrous histiocytoma. Modern pathology, 20(7), 749.

[17] Khan, J., Wei, J. S., Ringner, M., Saal, L. H., Ladanyi, M., Westermann, F., ... & Meltzer, P. S. (2001). Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks. Nature medicine, 7(6), 673.

[18] Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., & Mesirov, J. P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genomewide expression profiles. Proceedings of the National Academy of Sciences, 102(43), 15545-15550.

[19] Gordon, G. J., Jensen, R. V., Hsiao, L. L., Gullans, S. R., Blumenstock, J. E., Ramaswamy, S., ... & Bueno, R. (2002). Translation of microarray data into clinically relevant cancer diagnostic tests using gene expression ratios in lung cancer and mesothelioma. Cancer research, 62(17), 4963-4967.

[20] Tian, E., Zhan, F., Walker, R., Rasmusson, E., Ma, Y., Barlogie, B., & Shaughnessy Jr, J. D. (2003). The role of the Wnt-signaling antagonist DKK1 in the development of osteolytic lesions in multiple myeloma. New England Journal of Medicine, 349(26), 2483-2494.

[21] Güçkıran K., Cantürk İ., Özylmaz L. DNA Microarray Gene Expression Data Classification Using SVM, MLP, and RF with Feature Selection Methods Relief and LASSO. Süleyman Demirel Üniversitesi Fen Bilimleri Enstitüsü Dergisi. 2019; 23(1): 126-132.

[22] Arias-Michel, R., García-Torres, M., Schaerer, C. E., & Divina, F. (2015, September). Feature selection via approximated Markov blankets using the CFS method. In Data Mining with Industrial Applications (DMIA), 2015 International Workshop on (pp. 38-43). IEEE.

[23] Huertas, C., & Juarez-Ramirez, R. (2016). Automatic Threshold Search for Heat Map Based Feature Selection: A Cancer Dataset Analysis. World Academy of Science, Engineering and Technology, International Journal of Computer, Electrical, Automation, Control and Information Engineering, 10(7), 1341-1347.

[24] Huynh, P. H., Nguyen, V. H., & Do, T. N. (2018, December). Random ensemble oblique decision stumps for classifying gene expression data. In Proceedings of the Ninth International Symposium on Information and Communication Technology (pp. 137-144). ACM.

[25] Mundra, P. A., & Rajapakse, J. C. (2010). Gene and sample selection for cancer classification with support vectors based t-statistic. Neurocomputing, 73(13-15), 2353-2362.

[26] Le Thi, H.A., & Phan, D. N. (2017). DC programming and DCA for sparse Fisher linear discriminant analysis. Neural Computing and Applications, 28(9), 2809-2822.