

# Fully Adaptive Elastic-Net (FAElastic) For Gene Selection In High Dimensional Cancer Classification

Isah Aliyu Kargi, Norazlina bint Ismail, Ismail bin Mohammad

**Abstract:** Classification of cancer in high dimensional DNA microarray data establish a significant field of research. Though, because of the challenges face by higher dimensional data in selection of genes and classification, numerous penalized likelihood methods are unsuccessful in identifying a small subset of significant genes. To address this problem, the present study proposed and applied a Fully Adaptive Elastic-net (FAElastic) model to perform gene selection and estimation of gene coefficients simultaneously. The proposed techniques, FAElastic-net has been assessed in terms of AUC, number of genes selected, Sensitivity, Specificity and informedness. From the findings which was computed from colon cancer microarray data set, it was confirmed that FAElastic outperforms the other four techniques from the performance metrics which includes: (i) selected number of genes (ii)AUC (iii)Sensitivity and Specificity and (iv) informedness. In addition, FAElastic results can be used practically to other related data of high dimensionality for cancer classification. Thus, we can accomplish the efficiency of the proposed FAElastic-net technique in practice to the medical research area.

**Keywords:** Adaptive elastic net, Classification of cancer, Gene selection, Regularized logistic regression

## 1. INTRODUCTION

With recent development of high dimensional microarray data in genetic and molecular biology, the resultant sets of data clearly have a small size of sample with a higher dimension where the size of the sample is typically in the range of hundreds, with tens of thousands of number of genes[1], [2]. Therefore, high dimensionality of data is regarded as the major tasks of statistical modelling [3]–[5]. Multicollinearity and modelling overfitting are the major issues of high dimensional data when using methods of statistical classification and features selection. As a result model selection and classification techniques with microarray data becomes very difficult [6]–[8]. However, statistically, in DNA microarray cancer data, poor classification and overfitting may occur as a results of too many number of genes[9]. Due to the importance of these issues, an efficient and effective techniques are required to improve the classification accuracy and selecting small subset of genes appropriately. Furthermore, chosen significant genes will lead to an easy discovery and diagnosis of patients affected with cancer disease[10], [11]. Various statistical techniques have been adopted successfully in the field of bioinformatics for cancer classification. Logistic regression (LR) is one of the most effective approaches as it makes interpretations of the genes coefficient easily[9].

However, in high dimensional data scenario the LR is not suitable as a result of unavailability of the full rank of the design matrix in computation for classification of cancer [12]. Therefore, techniques of iterations like Newton Raphson's approach is not recommended [13]. Penalized logistic regression (PLR) has been applied successfully for

classification of cancer in data of higher dimensionality[14]–[17]. Least absolute shrinkage and selection operator (LASSO) proposed by [18] is the best commonly penalized techniques which imposes  $\ell_1 - norm$  penalization term into the loss function. As a result of  $\ell_1 - norm$  penalty, selection of variable can easily be achieved by LASSO by pushing some genes coefficient to exactly zero. Thus, for this motive, LASSO has gained acceptability in high dimensional cancer classification. Despite the ability of LASSO in performing feature/variable selection, it still has some limitations. For instance, in high dimensional microarray cancer classification, LASSO cannot choose variables more than the sample size (n), also in the presence of high correlation between the explanatory variables, LASSO can only choose one variable among the group of correlated variables [19]. To tackle this drawbacks of LASSO earlier stated, Elastic net penalty was proposed by Zou and Hastie [20] which consist of ridged penalty and LASSO penalty respectively. The idea behind the elastic net is to address the grouping effect of highly correlated genes in high dimensional cancer classification. In the same way like LASSO, elastic net does not enjoy the oracle property even though it has a better performance than LASSO. Additionally, Adaptive elastic net was proposed by zou and zhang[21] to address the effect of grouping and as well enjoying the oracle property. Furthermore, adaptive elastic net practically faces challenges in high dimensional cancer data by proposing maximum likelihood estimates(MLE) as the initial weight, and MLE is not suitable. Thus Adaptive elastic net is no longer appropriate. Employing elastic net estimates as the initial weight in the Adaptive elastic net was recommended by zou and zhang[21]. Though applying this weight in adaptive elastic net may not be appropriate for the following reasons: the  $\ell_2 - norm$  in elastic net forces genes of highly correlated coefficients to each other. But this property fails to discriminate highly correlated genes with significantly dissimilar magnitudes, particularly those with dissimilar signs [22]. This problem results in genes having significantly dissimilar magnitudes to be grouped together. To tackle this problem, in this study a new regularized method Fully Adaptive elastic net (FAElastic) estimator was proposed, which introduces an  $\ell_2 - norm$  to the original LASSO estimator. However, the objective is to add the  $\ell_2 - norm$  such that the effect of grouping does not deteriorates the

- Isah Aliyu Kargi 1Department of Mathematical Sciences, Universiti Teknologi Malaysia (UTM), 81310 Johor Bahru, Malaysia. E-mail: isaaliyukargi@gmail.com 2Department of Statistics, School of Applied Sciences Nuhu Bamalli Polytechnic Zaria Kaduna State.
- Norazlina bint Ismail 1Department of Mathematical Sciences, Universiti Teknologi Malaysia (UTM), 81310 Johor Bahru, Malaysia. E-mail: i-norazlina@utm.my
- Ismail bin Mohammad 1Department of Mathematical Sciences, Universiti Teknologi Malaysia (UTM), 81310 Johor Bahru, Malaysia. E-mail: immathematics@gmail.com

performance of the estimator to identify highly correlated genes with significantly dissimilar coefficient magnitudes.

## 2. Materials and Methods

### 2.1 Regularized logistic Regression

Logistic regression is a generalized linear model (GLM) used to predict binary categorical outcomes. The cost function used in logistic regression is defined as the sigmoid function which maps predicted values to probabilities, i.e., between 0 and 1. In cancer classification the outcomes are whether a patient's tumor is malignant or not based on a given data i.e., 0 for benign and 1 for malignancy. Let us denote the response variable for observation  $i$  by  $y_i \in (0,1)$ , where  $i = 1, 2, 3, \dots, n$ . Assume  $p$  genes with  $i$ -th gene vector represented by  $x_i = (x_{1,i}, x_{2,i}, x_{3,i}, \dots, x_{p,i})^T$ . Then the linear predictor is given by:

$$L(x_{1,i}\beta) = \beta_0 + \beta_1 x_{1,i} + \dots + \beta_p x_{p,i} \quad (1)$$

Given the linear predictor, the logit transformation is carried out as follows:

$$\log_e \left( \frac{p_i}{1-p_i} \right) = \beta_0 + \beta_1 x_{1,i} + \dots + \beta_p x_{p,i} \quad (2)$$

The probability of observation  $i$  belonging to the positive class, i.e., malignant, can now be calculated as follows:

$$P_i = P[y_i = 1 | x_i] = \frac{e^{\beta_0 + \beta_1 x_{1,i} + \dots + \beta_p x_{p,i}}}{1 + e^{\beta_0 + \beta_1 x_{1,i} + \dots + \beta_p x_{p,i}}} \\ = \frac{e^{x_i^T \beta}}{1 + e^{x_i^T \beta}} \quad (3)$$

Where,  $\beta = (\beta_0, \beta_1, \dots, \beta_p)^T$  is considered as the vector of unidentified gene coefficients. The log-likelihood function which defines the cost function in logistic regression is given as follows:

$$LL(\beta) = \sum_{i=1}^n \left[ y_i \log(p_i) + (1 - y_i) \log(1 - p_i) \right] \quad (4)$$

Cancer data is typically high dimensional data with  $p \gg n$ . To control the number of genes used in the predictions of the response variables, a non-negative regularization terms is added to the negative log-likelihood function in Eqn. (4). Some of the popular regularization terms in existing literature are discussed in [18], [20], [23], [24]. The most popular regularization term is the  $\ell_1$ -norm proposed in [18]. In regularized logistic regression (RLR) the  $\ell_1$ -norm is used as the penalty to constrain the negative log-likelihood function for gene selection and coefficient estimation, i.e.,  $-LL(\beta) + \lambda P(\beta)$ , where  $\lambda P(\beta)$  is used to regularize the coefficient estimates. Note that  $\lambda$  is the positive turning parameter used to control degree of shrinkage. For  $\lambda = 0$ , the regularization term is neglected and the RLR is reduced to logistic regression with the maximum likelihood estimator (MLE). However, a larger value of  $\lambda$  will have a higher influence on the estimated coefficients. The value of  $\lambda$  depends on the specific requirements of a model and is usually chosen

to minimize the mis-classification error rate by finding the optimum balance where the decrease in bias is equal to the increase in variance. Thus, the objective function for RLR is defined as:

$$\beta_{RLR} = \arg \min_{\beta} \left[ \sum_{i=1}^n \left\{ y_i \log(p_i) + (1 - y_i) \log(1 - p_i) \right\} + \lambda P(\beta) \right] \quad (5)$$

Without loss of generality, we assume that the gene vectors are standardized,

i.e.,  $\sum_{i=1}^n x_{ij} = 0$  and  $\frac{1}{n} \sum_{i=1}^n x_{ij}^2 = 1, \forall j \in \{1, 2, 3, \dots, p\}$ . One of the popular RLR techniques in literature is LASSO, the objective function for LASSO is given as follows:

$$\beta_{LASSO} = \arg \min_{\beta} \left[ \sum_{i=1}^n \left\{ y_i \log(p_i) + (1 - y_i) \log(1 - p_i) \right\} + \lambda \sum_{j=1}^p |\beta_j| \right] \quad (6)$$

Note that the value of turning parameter  $\lambda$  in LASSO has high impact on the coefficient estimates. Even a small value of  $\lambda$  may result in some coefficients reducing to absolute zero. This shows that LASSO has the ability to select some features while reducing the coefficients of the others zero. Therefore, LASSO has been widely used for feature selection in different application domains [25]–[27]. In practice the value of  $\lambda$  is usually selected using procedure of cross-validation to identify the  $\lambda$  which results in the minimum mis-classification error. The coordinate descent algorithm can be used to efficiently solve Eqn. (6) [23], [28], [29].

To address the two shortcomings of LASSO, discussed in section 1 another regularization method was proposed in [20]. To enable selecting more genes than the sample size and consider grouping in genes, elastic net introduces the  $\ell_2$ -norm to LASSO. The combination of  $\ell_1$  and  $\ell_2$  regularization improves the performance of genes selection and classification accuracy when genes are highly correlated and independent [7], [30]. RLR using elastic net is modelled using the following:

$$\beta_{Elastic} = \arg \min_{\beta} \left[ \sum_{i=1}^n \left\{ y_i \log(p_i) + (1 - y_i) \log(1 - p_i) \right\} + \lambda \left( (1 - \alpha) \sum_{j=1}^p |\beta_j| + \alpha \sum_{j=1}^p \beta_j^2 \right) \right] \quad (7)$$

Eqn. (7) shows that the elastic net estimator employs two turning parameters, i.e.,  $\lambda$  and  $\alpha$ . Note that an optimal combination of  $\lambda$  and  $\alpha$  is required for a good estimator. Cross validation can be used to find the optimal  $\lambda$  for a fixed  $\alpha$ .

### 2.2 Regularized Adaptive logistic regression

It has been shown in (2) that the LASSO treats all the genes coefficients equally leading to a bias as well as violating the oracle property. To solve this issue, Adaptive LASSO was proposed in [31]. Adaptive LASSO introduces adaptive weights for the coefficient into the penalty, i.e., we can prioritize various coefficients differently. Thus higher priority can be given to smaller coefficients, while lower priority can be given to larger coefficients, to achieve a balance and reduce the selection bias in the estimator. Adaptive LASSO can be modelled using the following objectives function:

$$\beta_{ALASSO} = \arg \min_{\beta} \left[ \sum_{i=1}^n \left\{ y_i \log(p_i) + (1 - y_i) \log(1 - p_i) \right\} + \lambda \sum_{j=1}^p w_j |\beta_j| \right] \quad (8)$$

Where  $w_j$  is the weight for the  $j$ -th coefficient and  $w_j = (\beta_j)^{-\gamma}$ , such that  $\gamma > 0$  and  $\beta$  is a root  $n$ -consistent initial value. These conditions ensure that the estimates converge to their true value with  $O(n^{-1/2})$  [14]. Adaptive LASSO addresses the issue of oracle property, however, the grouping effect problem of LASSO is inherited by adaptive LASSO as well. To achieve the grouping effect and improve the accuracy of LASSO, Elastic net introduces  $\ell_2$ -norm to LASSO as in [28]. However, elastic net does not satisfy the oracle property [21]. Therefore, adaptive elastic net was proposed in [21], [32] (13,4) as follows:

$$\beta_{AElastic} = \arg \min_{\beta} \left[ \sum_{i=1}^n \{y_i \log(p_i) + (1-y_i) \log(1-p_i)\} + \lambda \left( (1-\alpha) \sum_{j=1}^p w_j |\beta_j| + \alpha \sum_{j=1}^p \beta_j^2 \right) \right] \quad (9)$$

Adaptive elastic net improves over LASSO by achieving better selection consistency, grouping effect and oracle property [31]. Eqn.(9) can be solved using an augmented vector  $y^* = \begin{pmatrix} y \\ \alpha \end{pmatrix}$  as follows [20]:

$$\beta_{AElastic}^* = \arg \min_{\beta} \left[ \sum_{i=1}^n \left\{ y_i^* \log(p_i) + (1-y_i^*) \log(1-p_i) \right\} + \lambda \left( (1-\alpha) \sum_{j=1}^p w_j |\beta_j| + \alpha \sum_{j=1}^p \beta_j^2 \right) \right] \quad (10)$$

### 2.3. The Proposed Method

Adaptive elastic net enjoys the oracle property as well as achieves the grouping effect and selection consistency. However, the performance of the estimator is sensitive to the initial weight. Existing works on adaptive elastic net have proposed the use of ordinary least squares (OLS), MLE and elastic net for the estimation of the initial weights. However, OLS and MLE, may be suitable for low dimensional data, but for high dimensional cancer classification they are not appropriate [21], [32]. Similarly, using elastic net as the initial weight may also not be appropriate because of the following reasons: the  $\ell_2$ -norm in elastic net pushes the coefficients of highly correlated genes close to each other. However, this property fails to distinguish genes that are highly correlated with significantly dissimilar magnitudes, particularly those with dissimilar signs [22]. This issue results in genes having significantly different magnitudes to be grouped together. To solve this issue, in this paper we proposed an FAElastic estimator which introduces an  $\ell_2$ -norm to the original LASSO estimator. However, the objective is to add the  $\ell_2$ -norm in such a way that the grouping effect does not deteriorate the performance of the estimator to identify highly correlated genes with significantly different coefficient magnitudes

**The objective function for FAElastic is given as follows:**

$$\beta_{FAElastic} = \arg \min_{\beta} \left[ \sum_{i=1}^n \{y_i \log(p_i) + (1-y_i) \log(1-p_i)\} + \lambda \left( (1-\alpha) \sum_{j=1}^p w_j |\beta_j| + \alpha \sum_{j=1}^p \beta_j^2 \right) \right] \quad (11)$$

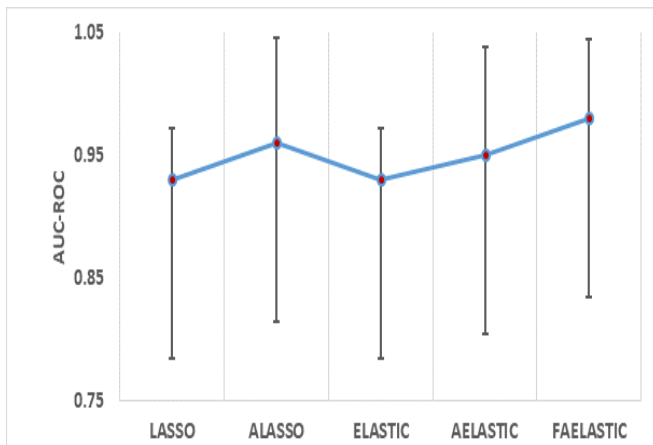
Note that in adaptive elastic net, the adaptive weights are only applied to the  $\ell_1$ -norm. However, in FAElastic we apply adaptive weight to the  $\ell_1$  as well as  $\ell_2$ -norms. To achieve the grouping effect, the turning parameter  $\alpha$  is tilted towards the  $\ell_2$ -norms while the initial weights are selected such that the coefficient of highly correlated genes are treated in an exclusive

manner, i.e., we used the LASSO estimator to generate the initial weights. Thus, using a value of  $\alpha > 0.5$  causes the estimator to maintain the grouping effect. Similarly, using LASSO estimator for the initial weights we achieve shrinkage but it does not cause the problems inherent to the LASSO estimator such as gene selection consistency because of the higher  $\alpha$  value.

### 3. RESULTS AND DISCUSSION

To evaluate the effectiveness of the proposed estimator, the DNA microarray data set for colon cancer was used [33]. The data set contains data for 40 tumor and 22 normal colon tissue for 6500 human genes obtained with Affymetrix oligonucleotide array. After data cleaning a total of 62 observations 2000 genes, i.e.,  $n = 60$  and  $P = 2000$  was used in this paper. The data set was divided into a training and testing data set. The training data set was randomly generated with 70% of the original data while the remaining 30% of the data was used as for testing. To ensure a fair comparison with the existing techniques, the training and test data partition was repeated 50 times. Moreover, the 10-fold cross validation (CV) method was used to find the optimal value of  $\lambda$ . To find the optimal value for  $\alpha$  we employed a turning Grid where the value of  $\alpha$  is incremented in steps and the resulting performance metric is recorded for each iteration. Then the value of  $\alpha$  which produces the best results is selected for evaluating an estimator. The open source software R was used to carry out all the computation in this paper. The performance metric used in this paper include the number of selected genes (#G), area under the receiver operating characteristics curve (AUC), sensitivity (SEN), specificity (SPEC), and informedness (IF). It can be seen from table 2 that our proposed method, FAElastic, has a better performance compared to other approaches in terms of number of selected genes, for instance, FAElastic selected 7 genes compared to 12,13,18 & 19 for LASSO, ALASSO, Elastic net and adaptive Elastic net respectively. Similarly, for classification precision in table 2 using AUC for colon cancer data, our proposed method FAElastic has a classification accuracy of 0.983, which is better than 0.930, 0.960, 0.935, 0.950 obtained by LASSO, ALASSO, Elastic net and Adaptive Elastic net respectively, which showed the best classification capability of the FAElastic than the other four techniques. We can also observe from table 1 that FAElastic outperforms the other four techniques from the performance metrics of Specificity and sensitivity. FAElastic and ALASSO are considered best methods with the highest sensitivity of 100% accuracy for colon cancer, this might be as a result of additional weight that is been added and oracle property enjoyed by both FAElastic and ALASSO. However, the other three methods have a lower sensitivity of 83.3% for LASSO and Elastic net and 94.1% for Adaptive Elastic net. This showed that FAElastic has significantly succeeded in classifying patients that have colon cancer with a sure probability value of 1.00 respectively. Furthermore, the outcomes of specificity in table 2 which display the likelihood of a Regularized adaptive logistic regression technique in classifying patients who are normal. The proposed method FAElastic, is superior than the other methods for the colon data set in terms of Specificity. For instance, FAElastic has the highest probability value of 0.906 in classifying the normal patients compared to 0.893, 0.844 and 0.852 for LASSO, ALASSO, Elastic net and Adaptive elastic net correspondingly. On the other hand, the results for informedness indicated the

superiority of our proposed technique. For example, FAElastic-net has largest IF of 0.906 in assessing predictive models for an imbalance data compared to the other four methods.



**Figure 1:** performance comparison of FAElastic with existing methods on the basis of average AUC: 95% confidence interval.

The confusion matrix for binary classifier can be represented as shown in table1. Sensitivity measures the proportion of the true positive that are correctly classified as such. For example, the percentage of patients with colon cancer that are correctly classified as having colon cancer by a classifier. Sensitivity, can be calculated from the table as follows:

$$Sen = \frac{TP}{T_{Positive}} \quad (12)$$

Similarly, Specificity measures the proportion for true negative that are correctly classified as such. For example, the percentage of patients that do not have colon cancer are correctly classified. Specificity can be calculated from the table 1 as follows:

$$Spec = \frac{TN}{T_{Negative}} \quad (13)$$

**Table 1:** Confusion matrix

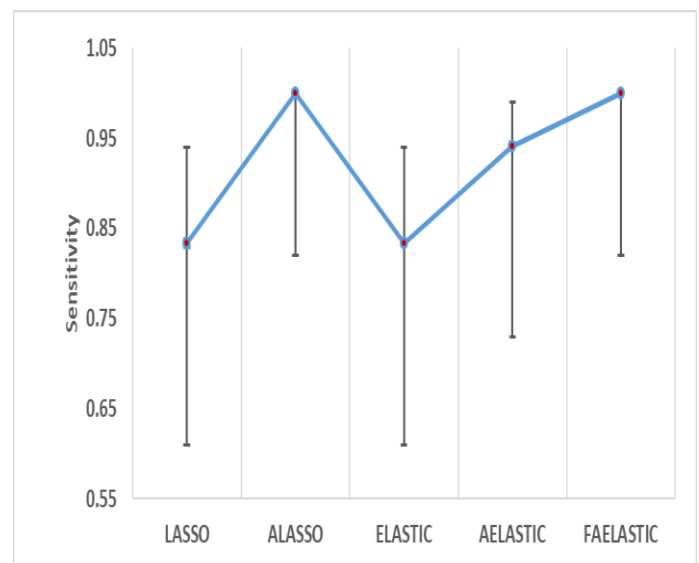
		True Condition		Total
		Malignant	Benign	
Predicted Condition	Malignant	True Positive (TP)	False Positive (FP)	$\hat{T}_{Positive}$
	Benign	False Negative (FN)	True Negative (TN)	$\hat{T}_{Negative}$
Total		$T_{Positive}$	$T_{Negative}$	T

For imbalanced data such as cancer data, classification accuracy and F1 are considered to be highly biased

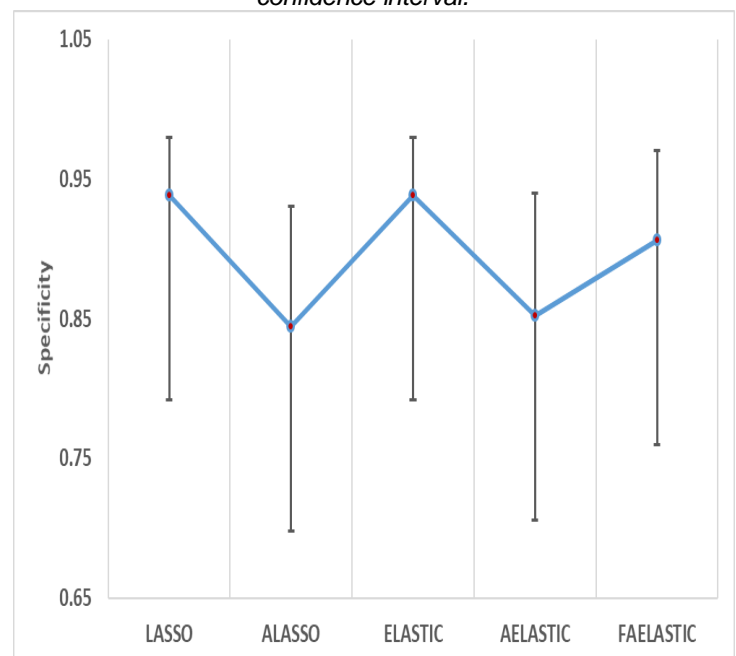
metrics[34]–[36]. Therefore, in this paper we use an alternate metric called informedness (IF). The superiority of IF in assessing predictive models using imbalanced data is shown in [37]. A higher value of IF represents better predictive power with a perfect model having a value of 1. We compare the performance of FAElastic estimator with LASSO [18], ALASSO [31], Elastic net [20] and Adaptive elastic net[21]). Table 2 present the results for different estimators. We observe that FAElastic outperformed the other estimators in all the criteria.

**Table 2:** Classification performance results on colon data.

Method	#G	AUC	Sen.	Spec.	IF
LASSO	13	0.930	0.833	0.893	0.771
ALASSO	12	0.960	1.00	0.844	0.843
Elastic Net	19	0.935	0.833	0.893	0.771
Adaptive Elastic Net	18	0.950	0.941	0.852	0.793
FAElastic	7	0.983	1.00	0.906	0.906



**Figure 2:** Performance comparison of FAElastic with existing method on the basis of average sensitivity: 95% confidence interval.



**Figure 3:** Performance comparison of FAElastic with existing methods on the basis of average specificity:95% confidence intervals.

Similarly, Figure 1,2 and 3 shows the 95% confidence interval for the mean of AUC, Sen., and Spec., respectively, on the testing data set. We can observe clearly that FAElastic outperforms the other classifiers. We have also seen that FAElastic results in the least variance for the performance metrics over the 50 partitions.

#### 4 CONCLUSION

Classification of cancer in high dimensional DNA microarray data establish a significant field of research. Though, because of the challenges face by higher dimensional data in selection of genes and classification, numerous penalized likelihood methods are unsuccessful in identifying a small subset of significant genes. To address this problem, the present study proposed and applied an FAElastic model to perform gene selection and estimation of gene coefficients simultaneously. The proposed techniques, FAElastic has been assessed in terms of AUC, number of genes selected, Sensitivity, Specificity and informedness. From the findings which was computed from colon cancer microarray data set, it was confirmed that FAElastic outperforms the other four techniques in terms of (i) selected number of genes (ii)AUC (iii)Sensitivity and Specificity and (iv) informedness. Generally, the results established the detail that FAElastic is a very viable technique that can analyse DNA microarray cancer data accurately. In addition, the proposed FAElastic results can be applied practically to other related high dimensional data for cancer classification. Thus, we can accomplish the efficiency of the proposed FAElastic technique in practice to the medical research area.

#### REFERENCES

- [1] N. A. Al-Thanoon, O. S. Qasim, and Z. Y. Algamal, "Tuning parameter estimation in SCAD-support vector machine using firefly algorithm with application in gene selection and cancer classification," *Comput. Biol. Med.*, vol. 103, no. August, pp. 262–268, 2018.
- [2] Z. Yahya, R. Alhamzawi, H. Taha, and M. Ali, "Gene selection for microarray gene expression classification using Bayesian Lasso quantile regression," *Comput. Biol. Med.*, vol. 97, no. April, pp. 145–152, 2018.
- [3] I. M. Johnstone and D. M. Titterton, "Statistical challenges of high-dimensional data," *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.*, vol. 367, no. 1906, pp. 4237–4253, 2009.
- [4] S. Ma, R. Fildes, and T. Huang, "Demand forecasting with high dimensional data: The case of SKU retail sales forecasting with intra- and inter-category promotional information," *Eur. J. Oper. Res.*, vol. 249, no. 1, pp. 245–257, 2016.
- [5] G. Aneiros, R. Cao, R. Fraiman, C. Genest, and P. Vieu, "Recent advances in functional data analysis and high-dimensional statistics," *J. Multivar. Anal.*, vol. 170, pp. 3–9, 2019.
- [6] L. Wang, H. Cheng, Z. Liu, and C. Zhu, "A robust elastic net approach for feature learning," *J. Vis. Commun. Image Represent.*, vol. 25, no. 2, pp. 313–321, 2014.
- [7] M. Ijaz, Z. Asghar, and A. Gul, "Computation Ensemble of penalized logistic models for classification of high-dimensional data," *Commun. Stat. - Simul. Comput.*, vol. 0, no. 0, pp. 1–17, 2019.
- [8] T. Basu and J. Einbeck, "Binary Credal Classification under Sparsity Constraints," no. January, 2020.
- [9] Y. Liang et al., "Sparse logistic regression with a L1/2 penalty for gene selection in cancer classification," *BMC Bioinformatics*, vol. 14, no. 1, pp. 1–12, 2013.
- [10] K. Chen, K. Wang, K. Wang, and M. Angelia, "Applying particle swarm optimization-based decision tree classifier for cancer classification on gene expression data," *Appl. Soft Comput. J.*, vol. 24, pp. 773–780, 2014.
- [11] A. Hossain, S. Muhammad, S. Islam, J. M. W. Quinn, and F. Huq, "Machine learning and bioinformatics models to identify gene expression patterns of ovarian cancer associated with disease progression and mortality," *J. Biomed. Inform.*, vol. 100, no. February, p. 103313, 2019.
- [12] P. T. Vu, A. A. Szpiro, and N. Simon, "Spatial Matrix Completion for Spatially-Misaligned and High-Dimensional Air Pollution Data," pp. 1–20, 2020.
- [13] C. Bielza, V. Robles, and P. Larrañaga, "Expert Systems with Applications Regularized logistic regression without a penalty term: An application to cancer classification with microarray data," *Expert Syst. Appl.*, vol. 38, no. 5, pp. 5110–5118, 2011.
- [14] Z. Y. Algamal and M. H. Lee, "Penalized logistic regression with the adaptive LASSO for gene selection in high-dimensional cancer classification," *Expert Syst. Appl.*, vol. 42, no. 23, pp. 9326–9332, 2015.
- [15] J. Fan et al., "Feature Augmentation via Nonparametrics and Selection ( FANS ) in High Dimensional Classification Feature Augmentation via Nonparametrics and Selection ( FANS ) in High Dimensional Classification \*," vol. 1459, no. September, 2015.
- [16] H. Huang, "Large dimensional analysis of general margin based classification methods," pp. 1–28.
- [17] A. R. Patil, "Combination of Ensembles of Regularized Regression Models with Resampling-Based Lasso Feature Selection in High Dimensional Data," 2020.
- [18] R. Tibshirani, "Regression Shrinkage and Selection via the Lasso Robert Tibshirani," *J. R. Stat. Soc. Ser. B*, vol. 58, no. 1, pp. 267–288, 1996.
- [19] J. Mandozzi, P. Bühlmann, J. Mandozzi, and B. Peter, "Hierarchical Testing in the High-Dimensional Setting with Correlated Variables Hierarchical Testing in the High-Dimensional Setting with Correlated Variables," vol. 1459, no. March, 2016.
- [20] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," pp. 301–320, 2005.
- [21] H. Z. A. H. H. Zhang, "ON THE ADAPTIVE ELASTIC-NET WITH A DIVERGING NUMBER OF PARAMETERS," vol. 37, no. 4, pp. 1733–1751, 2009.
- [22] J. Wang. S., Nan, B., Rosset, S., Zhu, "Random lasso," vol. 5, no. 1, pp. 468–485, 2011.
- [23] Z. Liu et al., "Sparse logistic regression with Lp penalty for biomarker identification," *Stat. Appl. Genet. Mol. Biol.*, vol. 6, no. 1, 2007.
- [24] N. Computing, J. Li, Y. Jia, and Z. Zhi-hua, "Partly adaptive elastic net and its application to microarray classification Partly adaptive elastic net and its application to microarray classification," no. May, 2012.
- [25] Z. Y. Algamal, R. Alhamzawi, and H. T. Mohammad Ali, "Gene selection for microarray gene expression classification using Bayesian Lasso quantile regression," *Comput. Biol. Med.*, vol. 97, no. April, pp. 145–152, 2018.

- [26] S. Kwon, S. Lee, and Y. Kim, "Moderately clipped LASSO," *Comput. Stat. Data Anal.*, vol. 92, pp. 53–67, 2015.
- [27] W. Rejchel, "Neurocomputing Oracle inequalities for ranking and U -processes with Lasso penalty," vol. 239, pp. 214–222, 2017.
- [28] M. Y. and T. H. Park, "L 1 -regularization path algorithm for generalized linear models," pp. 659–677, 2007.
- [29] J. Friedman, T. Hastie, and R. Tibshirani, "Regularization Paths for Generalized Linear Models via Coordinate Descent," vol. 33, no. 1, pp. 1–20, 2010.
- [30] X. Liu, S. Wang, H. Zhang, H. Zhang, Z. Yang, and Y. Liang, "Novel regularization method for biomarker selection and cancer classification," *IEEE/ACM Trans. Comput. Biol. Bioinforma.*, vol. PP, no. c, p. 1, 2019.
- [31] H. Zou, "The adaptive lasso and its oracle properties," *J. Am. Stat. Assoc.*, vol. 101, no. 476, pp. 1418–1429, 2006.
- [32] S. Ghosh, "On the grouped selection and model complexity of the adaptive," pp. 451–462, 2011.
- [33] U. A. Alon, N. B. Arkai, D. A. N. Otterman, K. G. Ish, S. Y. Barra, and D. M. Ack, "Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays," vol. 96, no. June, pp. 6745–6750, 1999.
- [34] H. Yin, "An Empirical Study on Preprocessing High-dimensional Class-imbalanced Data for Classification," pp. 1314–1319, 2015.
- [35] A. Luque, A. Carrasco, A. Martín, and A. De, "The impact of class imbalance in classification performance metrics based on the binary confusion matrix," *Pattern Recognit.*, vol. 91, pp. 216–231, 2019.
- [36] G. Wang, J. Yuen, C. Teoh, J. Lu, and K. Sze, "Least squares support vector machines with fast leave - one - out AUC optimization on imbalanced prostate cancer data," *Int. J. Mach. Learn. Cybern.*, no. 0123456789, 2020.
- [37] D. M.W, "Archived at the Flinders Academic Commons : This is the author ' s post-print version of this article . The published article is available at: © 2011 Bioinfo Publications Please note that any alterations made during the publishing process may not appear," vol. 2, no. 1, pp. 37–63, 2011.