

Maximal Covariance Complexity-Based Penalized Likelihood Method in High Dimensional Data

Isah Aliyu Kargi, Norazlina Bint Ismail, Ismail Bin Mohamad



Abstract: Classification of cancer and selection of genes is one of the most important application of DNA microarray data. As a result of the higher dimensionality of microarray data, classification and selection of gene techniques are frequently employed to support the professional systems in the diagnosing ability of cancer with higher precision in classification. Least absolute shrinkage and selection operator (LASSO) is one of the most popular method for cancer classification and gene selection in high dimensional data. However, Lasso has limitations of being biased and cannot select variables more than the sample size (n) in gene selection and classification of high dimensional microarray data. To address this problems, LASSO-CIF was proposed using scale invariant measure of maximal information complexity of covariance matrix Σ denoted $C_1(\Sigma)$ with weight modifications as data-adaptive alternative to the fairly arbitrary choice of the regularization term (λ_1) in the least absolute shrinkage and selection operator (LASSO). The results indicated the effectiveness of the proposed method LASSO-CIF over the classical LASSO. The evaluation criteria result shows that the proposed method, LASSO-CIF has a better performance in terms of AUC and number of genes selected

Keywords: Lasso, Maximal Complexity, Information Measure, Theoretic Measure, Penalized Likelihood Method, Scale-Invariant Complexity.

I. INTRODUCTION

With recent development of high dimensional microarray data in genetic and molecular biology, the resultant datasets clearly have a small size of sample with a higher dimension where the size of the sample is typically in the range of hundreds, whereas the number of genes is in tens of thousands[1], [2]. The success of any statistical method in high dimensional data rely on the pre-determination of dissimilar features[3]. The aim of feature selection and dimension reduction is identifying the least possible but most significant subset. As a result, various feature selection approach have been proposed in the literature[3]–[8]. Maximal Covariance Complexity-Based Penalized Likelihood Methods (MCBPM) had proven the capability among the outstanding system of expert classifiers in the

classification of high dimensional data [9]–[13]. This study aimed to compare Classical LASSO performance in gene selection and classification with a proposed method called LASSO-CIF method in classification of high dimensional data. Studies on classification of high dimensional microarray data have proposed different types of classical LASSO for variable selection and classification[13], [14].

Studies such as [15]–[19], proposed a new algorithm built on the Gauss–Seidel approach to solve penalized logistic regression (PLR) by usage in selection of gene and classification of microarray cancer data. J. Zhu, [16] trailed with alternative PLR proposals as a substitute for supporting vector machine in the classification of microarray cancer data considering the probability estimation. Furthermore, Shen and Tan,[17] come up with a combination of three dimension reduction approach, partial least square, singular value decomposition and PLR to improve the speed of computation and classification precision. A combination of ridge PLR and partial least squares was equally proposed by Fort and Lambert-Lacroix,[20], using prostate, colon and leukemia dataset to illustrate the classification performance. Kim, Kwon, and Heun Song,[21] proposed an extension of PLR to classify multi-class microarrays cancer data. A combination of Bayesian regularization and PLR in the selection of genes for classification of cancer data was proposed by Regularization, and Talbot [18]. A new method combining the non-convex penalty with PLR in the classification of cancer was developed by Liu *et al.*, [22], [23].

Additionally, Algamal and Lee,[24] presents a combination of Adaptive LASSO and PLR model for classification of gene in microarray data. Novel PLR method built on the advancement of regression coefficient with an estimate on distribution algorithm was developed by Zaragoza, *et al.* [25]. The major contribution of their work is resolving the penalization term during the selection of genes. Chen *et al.*[26] improves the algorithm of GLMNET for L_1 penalty to resolve some of the glmnet hypothetical execution problems. An investigation of the selection of gene and classification of cancer data with a new PLR having an $\ell_{1/2}$ penalty was developed by Bootkrajang and Kabán, [27]. Bayesian regularization was used to identify mislabeled arrays with PLR [27]. A new algorithm to resolve penalized LASSO with multinomial logistic regression (MLR) to classify multi-class gene data was proposed by Vincent and Hansen,[28].

The general property of a statistical model is complexity which is mostly sovereign of probabilistic specification, specific content or structure of the models. From the literature, the theory of complexity has been applied in various contexts.

Revised Manuscript Received on June 12, 2020.

* Correspondence Author

Isah Aliyu Kargi*, Department of Mathematics, Faculty of Science, Universiti Teknologi Malaysia 81310 UTM Skudai, Johor, Malaysia. 2 Department of Mathematics and statistics Nuhu Bamalli Polytechnic p.m.b 1061, Zaria Correspondence address: isaaliyukargi@gmail.com

Norazlina Bint Ismail, Department of Mathematics, Faculty of Science, Universiti Teknologi Malaysia 81310 UTM Skudai, Johor, Malaysia.

Ismail Bin Mohamad, Department of Mathematics, Faculty of Science, Universiti Teknologi Malaysia 81310 UTM Skudai, Johor, Malaysia.

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

According to Legg and Hutter, [29], [30] there is no precise definition of complexity in statistics, since the notation is indefinable.

In multivariate statistical data mining and modelling with a higher dimensional data matrix X of size $(n \times p)$, frequently p -variables interrelate in some manner and some variables (or sets of variables) can have significant effect to the other variables [29]. The interaction of These variables and their dependencies are reflected upon one another. Thus for a random vector, complexity can be defined as a measure of the degree of interdependency among the entire system and a simple enumerative structure of its parts or subsystems [11], [29], [31], [32].The more the dependency or interaction among the system the higher the complexity. This implies that a higher degree of complexity lead to a higher computation rates in statistical data processing [33]–[36]. Information theory is, therefore, suitable to evaluate the dependency or non-independency and as well measure the complexity of set of variables.

In higher dimensional DNA microarray data, least absolute shrinkage and selection operator (LASSO) is one of the most popular method for cancer classification and gene selection. However, despite its capability in performing variable selection and classification, it is limited by not selecting variables more than the sample size (n) and cannot select group of correlated variables. To improve the selection and classification ability of LASSO, the study proposed to apply scale invariant measure of maximal information complexity of covariance matrix Σ denoted $C_1(\Sigma)$ with weight modification as a data-adaptive and an alternative to the fairly arbitrary choice of the regularization term λ_1 in the least absolute shrinkage and selection operator [37].

The rest of this paper is organized as follows: the methodology applied in this study is detailed in section 2. In section 3, explanation of the data set and data analysis was displayed. Finally, section 4 covers the conclusion.

II. MATERIALS AND METHODS

2.1 Information Measure of Dependence in High-Dimensions

Bozdogan,[32] define the joint density function of p -dimensional distribution $f(\mathbf{x}) = f(x_1, \dots, x_p)$ with a marginal density functions $f_j(x_j), j = 1, \dots, p$, for information measure of dependency among the random variable X_1, \dots, X_p as:

$$I(\mathbf{x}) = I(x_1, \dots, x_p) = E_f[\log \frac{f(x_1, \dots, x_p)}{f_1(x_1) \dots f_p(x_p)}] \quad (1)$$

$$= \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} \log \frac{f(x_1, \dots, x_p)}{f_1(x_1) \dots f_p(x_p)} dx_1 \dots dx_p, \quad (2)$$

where I represent Kullback-Leibler (KL), information divergence against independence, and is a measure of the likely dependence between the component variables[38], [39].The joint distribution $f(x_1, \dots, x_p)$ of the random variables to the product of their marginal distribution $f_1(x_1)f_2(x_2) \dots f_p(x_p)$ is the major contrast base on the hypothesis that they are distributed independently to the level that the joint distribution significantly varies from the distribution of the random variable base on the theory of

dependency. Hence, this is known as measure of multicollinearity or interdependence between variables.

The properties of the Information are:

- a. $I(\mathbf{x}) \geq 0$ that is, the predictable information measure is nonnegative.
- b. $I(\mathbf{x}) = 0$ if and only if $f(x_1, \dots, x_p) = f_1(x_1) \dots f_p(x_p)$ for each p -tuple (x_1, \dots, x_p) ,

that is if and only if the random variables $x_1, \dots, x_p \sim$ are equal to unity, and its algorithm is then zero. But if is not equal to zero, it indicates dependency.

From the study of Bozdogan, [32], the joint entropy of KL divergence is define by the following identities $I(\mathbf{x}) \equiv I(x_1, \dots, x_p) = \sum_{j=1}^p H(x_j) - H(x_1, \dots, x_p)$, (3)

where

$$H(x_j) = -E[\log f(x_j)] = - \int_{-\infty}^{+\infty} f(x_j) \log f(x_j) dx_j \quad (4)$$

is the marginal entropy, and

$$H(x_1, \dots, x_p) = -E[\log f(x)]$$

$$H(x_1, \dots, x_p) = \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} f(x_1, \dots, x_p) \log f(x_1, \dots, x_p) dx_1 \dots dx_p \quad (5)$$

From eqn. (1), the relation can be generalized as a solution to the interaction amongst any subset of variables.

2.2 Covariance Complexity of Information Theoretic Measure

Bozdogan, [32] and Kartal and Hamparsum, [40] define the information-theoretic measure of complexity of a multivariate distribution, $(\mathbf{x}) = f(x_1, \dots, x_p)$ be a multivariate Gaussian ,the density function is defined by:

$$f(\mathbf{x}) = f(x_1, \dots, x_p)$$

$$f(\mathbf{x}) = (2\pi)^{-\frac{p}{2}} |\Sigma|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} (\mathbf{x} - \mu)' \Sigma^{-1} (\mathbf{x} - \mu) \right\} \quad (6)$$

where

$$\mu = (\mu_1, \mu_2, \dots, \mu_p)', -\infty < \mu_j < \infty, j = 1, 2, \dots, p$$

and $\Sigma > 0$ (positive definite).

and it can be written:

$$\mathbf{x} \sim N_p(\mu, \Sigma) \quad (7)$$

Formally, joint entropy $H(\mathbf{x}) = H(x_1, \dots, x_p)$ from eqn(3) for this scenario in which $\mu = 0$ is given by:

$$H(\mathbf{x}) = H(x_1, \dots, x_p) = - \int_{R^p} f(\mathbf{x}) \log f(\mathbf{x}) dx$$

$$H(\mathbf{x}) = \int_{R^p} f(\mathbf{x}) \left[\frac{p}{2} \log(2\pi) |\Sigma| + \frac{1}{2} (\mathbf{x} - \mu)' \Sigma^{-1} (\mathbf{x} - \mu) \right] dx \quad (8)$$

$$H(\mathbf{x}) = \frac{p}{2} \log(2\pi) |\Sigma| + \frac{1}{2} \text{tr} \left[\int_{R^p} f(\mathbf{x}) \Sigma^{-1} (\mathbf{x} - \mu) (\mathbf{x} - \mu)' dx \right].$$

Then, since $E[\mathbf{xx}'] = \Sigma$, we have

$$H(\mathbf{x}) = H(x_1, \dots, x_p) = \frac{p}{2} \log(2\pi) + \frac{p}{2} + \frac{1}{2} \log|\Sigma|$$

$$H(\mathbf{x}) = \frac{p}{2} [\log(2\pi) + 1] + \frac{1}{2} \log|\Sigma|. \quad (9)$$

Similarly, the marginal entropy $H(x_j)$ is

$$H(x_j) = - \int_{-\infty}^{+\infty} f(x_j) \log f(x_j) dx_j$$

$$H(x_j) = \frac{1}{2} \log(2\pi) + \frac{1}{2} + \frac{1}{2} \log(\sigma_j^2), \quad j = 1, 2, \dots, p \quad (10)$$

Where σ_j^2 is the variability of the j^{th} variable.

2.3 Covariance Complexity

Bozdogan,[32] and Bozdogan and Howe,[41], defined complexity of a covariance matrix Σ for multivariate normal distribution as:

$$I(x_1, \dots, x_p) \equiv C_0(\Sigma) = \sum_{j=1}^p H(x_j) - H(x_1, \dots, x_p)$$

$$= \sum_{j=1}^p \left[\frac{1}{2} \log(2\pi) + \frac{1}{2} \log(\sigma_{ij}) + \frac{1}{2} \right] - \frac{p}{2} \log(2\pi) - \log|\Sigma| - \frac{p}{2}. \quad (11)$$

This reduces to

$$C_0(\Sigma) = \frac{1}{2} \sum_{j=1}^p \log \sigma_{ij} - \frac{1}{2} \log|\Sigma|$$

$$C_0(\Sigma) = \frac{1}{2} \log|Diag(\Sigma)| - \frac{1}{2} \log|\Sigma|, \quad (12)$$

Where $\sigma_{ij} \equiv \sigma_j^2$, is the discrepancy or variability of the j^{th} variable, and also the j^{th} diagonal element of Σ . Hence, under orthonormal transformation, the first term of eqn. (12) is not invariant. This indicated the outcome in eqn. (12) is ineffective to measure the quantity of complexity in the covariance matrix Σ , since:

- i. $C_0(\Sigma)$ depends on the coordinates of the original random variables x_1, \dots, x_p
- ii. From eqn (10), $C_0(\Sigma)$ would change under orthonormal transformation

2.4 Maximal Covariance Complexity

Generally, the maximal covariance complexity of a matrix Σ can be expressed as the sovereign of the coordinates of the novel chance variables x_1, \dots, x_p associated with the variances $\sigma_j^2, j = 1, 2, \dots, p$. Since $C_0(\Sigma)$ in eqn.(10) is coordinate dependence. Thus, to improve upon $C_0(\Sigma)$ in eqn.(10) Bozdogan,[32] present the maximum of complexity in Σ under orthogonal transformations. Therefore, the maximal information theoretic measure of complexity of a covariance matrix Σ of a multivariate Gaussian distribution is given by :

$$C_1(\Sigma) = \max_{\tau} C_0(\Sigma) = \max_{\tau} \{H(x_1) + \dots + H(x_p) - H(x_1, \dots, x_p)\}$$

$$\frac{p}{2} \log \left[\frac{tr(\Sigma)}{p} \right] - \frac{1}{2} \log|\Sigma|. \quad (13)$$

Based on $C_0(\Sigma)$ measure, various studies have employed using the maximal complexity information measure $C_1(\Sigma)$ as a penalty functional [29], [32], [42].

If we let $\lambda_1, \lambda_2, \dots, \lambda_p$ represent the eigenvalues of the covariance matrix Σ , then

$\bar{\lambda}_a = tr(\Sigma)/p = 1/p \sum_{j=1}^p \lambda_j$
is considered as the arithmetic mean of the eigenvalues, also
 $\bar{\lambda}_g = |\Sigma|^{1/p} = (\prod_{j=1}^p \lambda_j)^{1/p}$
be the geometric mean of the eigenvalues of Σ . Therefore, the complexity of Σ can be written as

$$C_1(\Sigma) = \frac{p}{2} \log \left(\frac{\bar{\lambda}_a}{\bar{\lambda}_g} \right). \quad (14)$$

The log ratio among the geometric mean and the arithmetic mean of the eigenvalues of Σ is called complexity [32]. It evaluates the difference between the eigenvalues of Σ and also consider determinant and trace as the two scalar measures of multivariate scatter into one single function. Similarly, Frobenius norm is another form of complexity techniques[43], [44] and is given by :

$$C_F(\Sigma) = \frac{1}{p} \|\Sigma\|^2 - \left(\frac{tr \Sigma}{p} \right)^2, \quad (15)$$

And the square of the frobenius norm of Σ is invariant under the orthogonal transformation and can be presented as: $\|\Sigma\|^2 = tr(\Sigma' \Sigma)$. However, in terms of singular values (or eigenvalues), $C_F(\Sigma)$ becomes:

$$C_F(\Sigma) = \frac{1}{p} \sum_{j=1}^p (\lambda_j - \bar{\lambda}_a)^2, \quad (16)$$

So, P represent the rank of Σ , λ_j is the j^{th} eigenvalues of $\Sigma > 0, j = 1, 2, \dots, p$ and $\bar{\lambda}_a$ is the eigenvalues of the arithmetic mean.

2.5 Scale-Invariant Complexity

The connection between Frobenius complexity $C_F(\Sigma)$ and maximal information complexity $C_1(\Sigma)$ is the fact that they are all second order equivalent [32], [45]. Hence, H. Bozdogan,[32] approximated $C_1(\Sigma)$ as the eigenvalues $\lambda_j, j = 1, 2, \dots, p$ by :

$$C_1(\Sigma) \equiv \frac{1}{4} \sum_{j=1}^p \left(\frac{\lambda_j - \bar{\lambda}_a}{\bar{\lambda}_a} \right)^2 \quad (17)$$

Thus, we can relate $C_1(\Sigma)$ to the norm of Frobenius classification of complexity $C_F(\Sigma)$ of Σ by presenting $C_{1F}(\Sigma)$ which finally can be define as:

$$C_F(\Sigma) = \frac{p}{4} \frac{C_1(\Sigma)}{\left(\frac{tr(\Sigma)}{p} \right)^2} = \frac{p}{4} \frac{\|\Sigma\|^2 - \left(\frac{tr(\Sigma)}{p} \right)^2}{\left(\frac{tr(\Sigma)}{p} \right)^2} \quad (18)$$

$$C_F(\Sigma) = \frac{p}{4} \frac{\frac{1}{p} tr(\Sigma' \Sigma) - \left(\frac{tr(\Sigma)}{p} \right)^2}{\left(\frac{tr(\Sigma)}{p} \right)^2} \quad (19)$$

$C_F(\Sigma)$ in terms of the eigenvalues, becomes

$$C_{1F}(\Sigma) = \frac{p}{4} \frac{1}{p\bar{\lambda}_a^2} \sum_{j=1}^p (\lambda_j - \bar{\lambda}_a)^2 \quad (20)$$

$$= \frac{1}{4\bar{\lambda}_a^2} \sum_{j=1}^p (\lambda_j - \bar{\lambda}_a)^2. \quad (21)$$

Thus, $C_{1F}(\Sigma)$ is as the second order corresponding complexity measure to the novel $C_1(\Sigma)$ measure. Also, $C_{1F}(\Sigma)$ is term as the scale-invariant and $C_{1F}(\Sigma) \geq 0$ with $C_{1F}(\Sigma) = 0$ only when all $\lambda_j = \bar{\lambda}_a$. Furthermore, $C_{1F}(\Sigma)$ evaluate the relative variability in the eigenvalues before an absolute variability of the eigenvalues.

2.6 Application of $C_{1F}(\Sigma)$ in LASSO Shrinkage

The regression coefficient of the penalized linear regression using LASSO is given by:

$$\hat{\beta} = \underset{\text{subject to: } \|\beta\|_1 \leq t}{\operatorname{argmin}} \left\{ \frac{1}{2n} \sum_{i=1}^n (y_i - \sum_j \beta_j x_{ij})^2 \right\} \quad (22)$$

the tuning parameter t also called regularization or penalty parameter is usually computer-generated as grid of values. This implies that randomly generated t values will vary each time the LASSO procedure is repeated for the same data. As a result, the current study intuitively considered generating t values data-adaptively could make t arbitrary, data-specific and finally, corrects interdependency (or correlation) effects on LASSO procedure. Hence, in langrange form eqn. (22) can be written as:

$$\hat{\beta} = \underset{\text{subject to: } \|\beta\|_1 \leq t}{\operatorname{argmin}} \left\{ \frac{1}{2n} \sum_{i=1}^n (y_i - \sum_j \beta_j x_{ij})^2 + \lambda \|\beta\|_1 \right\} \quad (23)$$

$$\hat{\beta} = \underset{\text{subject to: } \|\beta\|_1 \leq t}{\operatorname{argmin}} \left\{ \frac{1}{2n} \|y - X\beta\|_2^2 + \lambda \|\beta\|_1 \right\} \quad (24)$$

The relation between eqn. (23) and eqn. (24) holds due to the duality and the KKT (Karush-Kuhn Tucker) conditions. Additionally, for every $t > 0$ there exists a $\lambda > 0$ such that both equations lead to the similar result [44], [46].

However, we notice that $C_{1F}(\Sigma)$ is a scaler whose value can be large depending on the structural complexity of Σ and can therefor make LASSO shrink all coefficients to zero. The current study will address his challenge by constructing $C_{1F}(\Sigma)$ for each dimension of Σ instead of the overall complexity of the entire Σ . Because is difficult to attach non statistical significance to differences in maximal covariance complexity scores when attempting to select the most appropriate model. Therefore, we computed the relative weights that can be interpreted as the probability that a given model due to a corresponding λ value is the most appropriate. These weights are mathematically computed as:

$$W_i = \frac{e^{-(C_{1Fi}(\Sigma) - \min(C_{1Fi}(\Sigma)))/2}}{\sum_{i=1}^K e^{-(C_{1Fi}(\Sigma) - \min(C_{1Fi}(\Sigma)))/2}} \quad (25)$$

where i indexes the K models evaluated, and $C_{1Fi}(\Sigma)$ denotes the maximal covariance complexity due inclusion of variable i to the K models.

III. DISCUSSION OF RESULTS

To Verify the efficiency of the suggested method, the classical LASSO algorithm in the *glmnet* package of the R software for statistical computing and graphics was modified. The classical LASSO algorithm and the modified algorithm were tested using a training set with 80% of the original size and a testing set with 20% over 50 Monte-Carlo cross-validation (MCCV) iterations with two microarray data sets. These data sets include the Leukemia data set of Golub *et al.*, [47] comprising expression levels of 7129 genes for 47 acute lymphoblastic leukemia (ALL) patients and 25 acute myeloid leukemia (AML) patients and the Prostate data set of Singh *et al.*, [48] comprising 12600 genes for 50 normal and 52 tumor patients, respectively.

Standard statistical classifier performance metrics including misclassification error rates (MER), correct classification rates (CCR), sensitivity (SEN), specificity (SPEC), positive predictive values (PPV), negative predictive values (NPV), balance accuracy (BA), G-means and area under the ROC curve (AUC), respectively, were estimated on the 20% test data set over 50 Monte-Carlo cross-validation (MCCV) iterations. The results are presented in tables one to tables ten (table 1-table 10) for the two data sets.

3.1 Leukemia Cancer Data Set results

3.1.1 Test performances:

Based on AUC error loss in Table 1, the proposed LASSO-C1F penalized logistic classifier consistently outperformed the classical LASSO penalized logistic classifier in terms of average CCR, SENs, SPECs, BAs, PPVs, NPVs, GM, AUC, respectively, over 50 Monte-Carlo cross-validation (50 MCCV) iterations. It is showed from these results that more than 29 genes selected by the classical LASSO may be important core relevant (biomarker) genes for discrimination of patients in terms of their leukemia cancer statuses. This inference is because with 35 non-zero genes, the proposed LASSO-C1F achieved better test prediction results. We can also observe that on deviance loss in Table 2, similar result was obtained which shows the efficiency of LASSO-C1F over the classical LASSO.

However, in Table 3, both the classical LASSO and proposed compete favorably well in terms of all average performance metrics to two places of decimal values. Though, the proposed LASSO selected additional 4 genes to achieve these classification performances. The proposed LASSO-C1F selects higher number of genes than the classical LASSO because it has the property that it selects not just one out of the group of correlated genes since its choice of shrinkage term is data-adaptive with consideration of covariance information complexity and interdependence among such genes.

Similarly, the results in Table 4, indicated that LASSO-C1F also favorably competes with classical LASSO in terms of true positive prediction rates (SENs), true negative prediction rates (SPECs), PPVs, and NPVs.

However, LASSO-C1F outperformed the classical LASSO in terms of CCRs, AUC and it identifies additional 6 core relevant genes for prediction of patient’s leukemia status which possibly were not identifiable by the classical LASSO. On the other hand, results for misclassification error measures in Table 5, LASSO-C1F penalized logistic regression classifier outperform classical LASSO logistic classifier on classification performances. It also identified additional 6 genes as core biomarkers for leukemia prediction in patients as it has consistently done under MAE, MSE, AUC and deviance error measures.

3.2 Prostate Cancer Data Set of Singh et al., 2002

3.2.1 Test performances:

Prostate cancer contains 12,600 genes for 52 prostate sample tumor and non-tumor tissues of 50. A subgroup of 5966 genes was modified in the classification [49]. The performance accuracy in the training set for the proposed LASSO-C1F method favorably competes with the classical LASSO in Table 6 but LASSO-C1F selects more relevant genes as it does for the leukemia cancer data. Also the results on the deviance error loss in Table 7, the evaluation criteria result shows that the proposed method, LASSO-C1F has a better performance in terms of CCRs, SENs, AUC (0.9486), (0.8904) and (95.31%) respectively. But similar performances are achieved by both methods in term of specificity. Furthermore, in terms of MAE loss in Table 8, both classical LASSO and LASSO-C1F compete favorably in terms of all classification performance metrics. Similarly, result from table 9 indicated that among the two methods

there is no anyone that outperform the other in all measure of accuracies.

Additionally, results from table 10 display the performance of the proposed method, LASSO-C1F and the classical LASSO, it can be seen from the result that the newly proposed method is superior than the classical LASSO in terms of CCRs, SENs, and AUC. But a similar result for the two methods in terms of SPECs for 50 MCCV Iterations with MER Loss for Shrinkage Parameter λ over 10-Fold CV respectively.

Although it is known in standard statistical learning and data mining literature that training performances will always underestimate test performances, training results are presented in figure 3.1 for both LASSO and LASSO-C1F logistic classifiers for leukemia cancer data set. The results prove the efficiency of LASSO-C1F over the classical LASSO both in classification precision and in selecting higher number of genes this might be as a result of the weight modifications as data-adaptive alternative to the fairly arbitrary choice of the regularization term which encourages grouping effect.

Similarly, Figure 3.2 demonstrate the training performance of the two techniques for prostate cancer datasets. Like leukemia data, figure 3.2 display the result of the proposed method, LASSO-C1F and the Classical LASSO and show a similar outcome in terms of CCRs, SENs, SPECs and AUC. However, LASSO-C1F selected higher number genes respectively this might be due to weight modifications as data-adaptive alternative to the fairly arbitrary choice of the regularization term.

Table 1: 50 MCCV Iterations with AUC Error Loss for Shrinkage Parameter λ over 10-Fold CV

Method	CCRs	SENs	SPECs	BAs	PPVs	NPVs	GM	AUC	# Genes
LASSO	0.9343	0.8574	0.9779	0.9176	0.9519	0.9264	0.9104	93.9168	29
LASSO-C1F	0.9486	0.8904	0.9795	0.9349	0.9619	0.9443	0.9307	95.31201	35

AUC: Area under the Curve

Table 2: 50 MCCV Iterations with Deviance Loss for Shrinkage Parameter λ over 10-Fold CV

Method	CCRs	SENs	SPECs	BAs	PPVs	NPVs	GM	AUC	# Genes
LASSO	0.9123	0.8024	0.9345	0.9026	0.9210	0.9250	0.9024	90.2145	29
LASSO-C1F	0.9524	0.8543	0.9536	0.9325	0.9450	0.9325	0.9416	93.4132	35

Table 3: 50 MCCV Iterations with MAE Loss for Shrinkage Parameter λ over 10-Fold CV

Method	CCRs	SENs	SPECs	BAs	PPVs	NPVs	GM	AUC	# Genes
LASSO	0.9429	0.8769	0.9795	0.9282	0.9619	0.9371	0.9229	94.9532	29
LASSO-C1F	0.9486	0.8896	0.9795	0.9345	0.9619	0.9447	0.9302	95.3303	33

MAE: Mean Absolute Error

Table 4: 50 MCCV Iterations with MSE Loss for Shrinkage Parameter λ over 10-Fold CV

Method	CCRs	SENs	SPECs	BAs	PPVs	NPVs	GM	AUC	# Genes
LASSO	0.9429	0.8769	0.9795	0.92818	0.9619	0.9371	0.9229	94.95317	29
LASSO-C1F	0.9500	0.8929	0.9795	0.9362	0.9619	0.9465	0.9321	95.41916	35

MSE: Mean Squared Error

Table 5: 50 MCCV Iterations with MER Loss for Shrinkage Parameter λ over 10-Fold CV

Method	CCRs	SENs	SPECs	BAs	PPVs	NPVs	GM	AUC	# Genes
LASSO	0.9271	0.8344	0.9761	0.9052	0.9418	0.9196	0.8957	93.0734	29
LASSO-C1F	0.9471	0.8871	0.9795	0.9333	0.9619	0.9425	0.9288	95.2231	35

MER: Misclassification Error Rate

Table 6: 50 MCCV Iterations with AUC Error Loss for Shrinkage Parameter λ over 10-Fold CV

Method	CCRs	SENs	SPECs	BAs	PPVs	NPVs	GM	AUC	# Genes
LASSO	0.9343	0.8574	0.9779	0.9176	0.9519	0.9264	0.9104	93.9168	29
LASSO-C1F	0.9486	0.8904	0.9795	0.9349	0.9619	0.9443	0.9307	95.31201	35

AUC: Area under the Curve

Table 7: 50 MCCV Iterations with Deviance Loss for Shrinkage Parameter λ over 10-Fold CV

Method	CCRs	SENs	SPECs	BAs	PPVs	NPVs	GM	AUC	# Genes
LASSO	0.9343	0.8574	0.9779	0.9176	0.9519	0.9264	0.9104	93.9168	29
LASSO-C1F	0.9486	0.8904	0.9795	0.9349	0.9619	0.9443	0.9307	95.3120	35

Table 8: 50 MCCV Iterations with MAE Loss for Shrinkage Parameter λ over 10-Fold CV

Method	CCRs	SENs	SPECs	BAs	PPVs	NPVs	GM	AUC	# Genes
LASSO	0.9429	0.8769	0.9795	0.9282	0.9619	0.9371	0.9229	94.9532	29
LASSO-C1F	0.9486	0.8896	0.9795	0.9345	0.9619	0.9447	0.9302	95.3303	33

MAE: Mean Absolute Error

Table 9: 50 MCCV Iterations with MSE Loss for Shrinkage Parameter λ over 10-Fold CV

Method	CCRs	SENs	SPECs	BAs	PPVs	NPVs	GM	AUC	# Genes
LASSO	0.9429	0.8769	0.9795	0.92818	0.9619	0.9371	0.9229	94.95317	29
LASSO-C1F	0.9500	0.8929	0.9795	0.9362	0.9619	0.9465	0.9321	95.41916	35

MSE: Mean Squared Error

Table 10: 50 MCCV Iterations with MER Loss for Shrinkage Parameter λ over 10-Fold CV

Method	CCRs	SENs	SPECs	BAs	PPVs	NPVs	GM	AUC	# Genes
LASSO	0.9271	0.8344	0.9761	0.9052	0.9418	0.9196	0.8957	93.0734	29
LASSO-C1F	0.9471	0.8871	0.9795	0.9333	0.9619	0.9425	0.9288	95.2231	35

MER: Misclassification Error Rate.

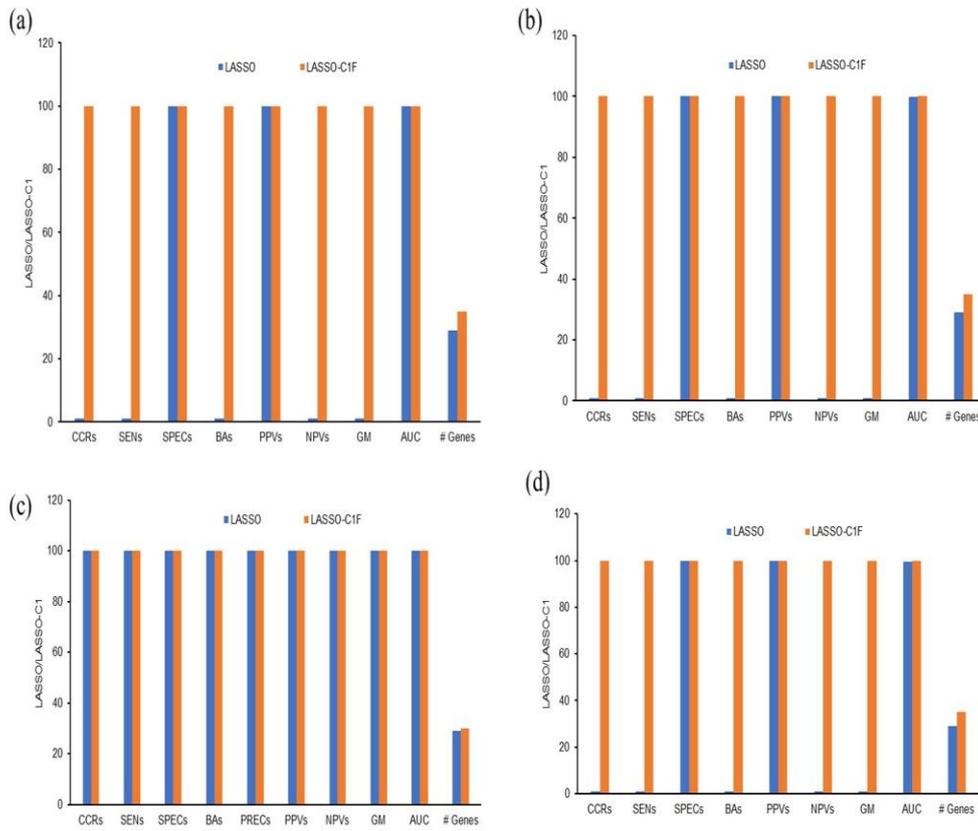


Figure 3.1. Results of training performance for leukemia data set.

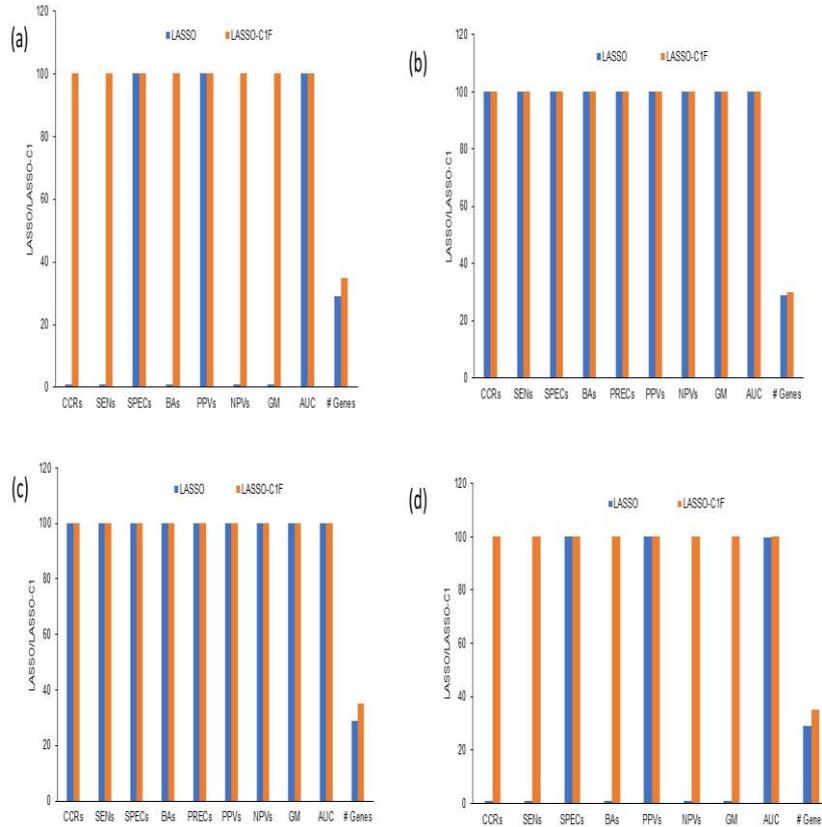


Figure 3.2: Results of training performance for Prostate data set.

IV. CONCLUSION

Overall, analysis of high dimensional DNA microarray data set establishes a very significant area of study in classification of cancer and gene selection. In the present paper we proposed LASSO-C1F that proved to more efficient in classifying high dimensional microarray data. Based on the revelation of this study, it can be inferred that:

- i. The efficiency of LASSO-C1F over the classical LASSO both in classification precision and reliability in genes selection is higher.
- ii. The performance accuracy in the training set for the proposed LASSO-C1F method favorably competes with the classical LASSO but LASSO-C1F selects more relevant genes as it does for the leukemia cancer data;
- iii. the evaluation criteria result shows that the proposed method, LASSO-C1F has a better performance in terms of CCRs, SENs, AUC (0.9486), (0.8904) and (95.31%) respectively;
- iv. In terms of MAE loss, both classical LASSO and LASSO-C1F compete favorably for all classification performance metrics;

Therefore, the newly proposed method can apply efficiently to analysed high dimensional microarray data for cancer classification and genes selection. LASSO-C1F can consistently perform selection of gene and address the consequence of grouping concurrently. Finally, the proposed method has shown the ability of choosing more correlated genes.

CONFLICT OF INTEREST

No conflict declares by the authors

ACKNOWLEDGMENTS

This research is financially supported by Tertiary education trust fund Nigeria (TETFUND) TETF/ES/POLY/ZARIA/ASTD/2017.

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. February, pp. 145–152, 2018.
3. S. Rakkeitinai, C. Lursinsap, C. Apornawan, and A. Mutirangura, "New feature selection for gene expression classification based on degree of class overlap in principal dimensions," *Comput. Biol. Med.*, vol. 64, pp. 292–298, 2015.
4. B. Yuzbasi, M. Arashi, and S. E. Ahmed, "Big Data Analysis Using Shrinkage Strategies," 2017.
5. L. S. Chen *et al.*, "Insights into Colon Cancer Etiology via a Regularized Approach to Gene Set Analysis of GWAS Data," *Am. J. Hum. Genet.*, vol. 86, no. 6, pp. 860–871, 2010.
6. S. A. Van de Geer and P. Bühlmann, "On the conditions used to prove oracle results for the lasso," *Electron. J. Stat.*, vol. 3, pp. 1360–1392, 2009.
7. C. Ding and C. Xing, "Journal of Complexity," *Coding Theory Cryptogr.*, vol. 32, pp. 20–39, 2004.
8. S. Zheng and W. Liu, "An experimental comparison of gene selection by Lasso and Dantzig selector for cancer classification," *Comput. Biol. Med.*, vol. 41, no. 11, pp. 1033–1040, 2011.

9. M. Figueiredo and R. Nowak, "An EM Algorithm for Wavelet-Based Image Restoration (random shift)," *Ieeexplore.Ieee.Org*, vol. 12, no. 8, pp. 906–916, 2003.
10. G. Wang and J. Qi, "Analysis of penalized likelihood image reconstruction for dynamic PET quantification," *IEEE Trans. Med. Imaging*, vol. 28, no. 4, pp. 608–620, 2009.
11. J. A. Howe and H. Bozdogan, "Regularized SVM Classification with a new Complexity-Driven Stochastic Optimizer," vol. 9, no. 2, pp. 216–230, 2016.
12. R. Wang *et al.*, "Toward Automated Calibration of Data Center Digital Twins: A Neural Surrogate Approach," 2020.
13. A. Afsar, A. Asif, and M. Arif, "crossmark," *Comput. Biol. Med.*, vol. 79, no. May, pp. 68–79, 2016.
14. E. Zintzaras and A. Kowald, "Forest classification trees and forest support vector machines algorithms : Demonstration using microarray data," *Comput. Biol. Med.*, vol. 40, no. 5, pp. 519–524, 2010.
15. S. K. Shevade and S. S. Keerthi, "A simple and efficient algorithm for gene selection using sparse logistic regression," *Bioinformatics*, vol. 19, no. 17, pp. 2246–2253, 2003.
16. J. Zhu, "Classification of gene microarrays by penalized logistic regression," no. July 2004, 2004.
17. L. Shen and E. C. Tan, "Dimension Reduction-Based Penalized Logistic Regression for Cancer Classification Using Microarray Data," vol. 2, no. 2, pp. 166–175, 2005.
18. B. Regularization, G. C. C. Å, and N. L. C. Talbot, "Gene Selection in Cancer Classification Using Sparse Logistic Regression with Gene expression Gene selection in cancer classification using sparse logistic regression with Bayesian regularization," no. January, 2006.
19. Z. Liu, "Statistical Applications in Genetics and Molecular Biology Sparse Logistic Regression with Lp Penalty for Biomarker Identification Sparse Logistic Regression with Lp Penalty for Biomarker Identification *," vol. 6, no. 1, 2007.
20. G. Fort and S. Lambert-Lacroix, "Classification using partial least squares with penalized logistic regression," *Bioinformatics*, vol. 21, no. 7, pp. 1104–1111, 2005.
21. Y. Kim, S. Kwon, and S. Heun Song, "Multiclass sparse logistic regression for classification of multiple cancer types using gene expression data," *Comput. Stat. Data Anal.*, vol. 51, no. 3, pp. 1643–1655, 2006.
22. Z. Liu *et al.*, "Sparse logistic regression with Lp penalty for biomarker identification," *Stat. Appl. Genet. Mol. Biol.*, vol. 6, no. 1, 2007.
23. N. Nidheesh, K. A. A. Nazeer, and P. M. Ameer, "An enhanced deterministic K-Means clustering algorithm for cancer subtype prediction from gene expression data," *Comput. Biol. Med.*, vol. 91, no. June, pp. 213–221, 2017.
24. Z. Y. Algamal and M. H. Lee, "Penalized logistic regression with the adaptive LASSO for gene selection in high-dimensional cancer classification," vol. 42, pp. 9326–9332, 2015.
25. J. H. Zaragoza, L. E. Sucar, E. F. Morales, C. Bielza, and P. Larranaga, "Bayesian chain classifiers for multidimensional classification," *IJCAI Int. Jt. Conf. Artif. Intell.*, pp. 2192–2197, 2011.
26. L. C. Chen *et al.*, "Human breast cancer cell metastasis is attenuated by lysyl oxidase inhibitors through down-regulation of focal adhesion kinase and the paxillin-signaling pathway," *Breast Cancer Res. Treat.*, vol. 134, no. 3, pp. 989–1004, 2012.
27. J. Bootkrajang and A. Kabán, "Classification of mislabelled microarrays using robust sparse logistic regression," *Bioinformatics*, vol. 29, no. 7, pp. 870–877, 2013.
28. M. Vincent and N. R. Hansen, "Sparse group lasso and high dimensional multinomial classification," vol. 71, pp. 771–786, 2014.
29. B. H., "Akaike Information Criterion and Recent Developments in Information Complexity," *J. Math. Psychol.*, vol. 44, no. 1, pp. 62–91, 2000.
30. S. Legg and M. Hutter, "Universal intelligence: A definition of machine intelligence," *Minds Mach.*, vol. 17, no. 4, pp. 391–444, 2007.
31. J. Teller, B. N. Unger, A. Kock, and H. G. Gemünden, "Formalization of project portfolio management: The moderating role of project portfolio complexity," *Int. J. Proj. Manag.*, vol. 30, no. 5, pp. 596–607, 2012.
32. H. Bozdogan, "A new class of information complexity (ICOMP) criteria with an application to customer profiling and segmentation," *İstanbul Üniversitesi İşletme Fakültesi Derg.*, vol. 39, no. 2, pp. 370–398, 2009.



33. F. J. Pulgar, F. Charte, A. J. Rivera, and M. J. Jesus, "Choosing the proper autoencoder for feature fusion based on data complexity and classifiers : Analysis , tips and guidelines," *Inf. Fusion*, vol. 54, no. July 2019, pp. 44–60, 2020.
34. T. Redel, "The two tontituntuudulhor amin," vol. 1, 2018.
35. J. Dean and S. Ghemawat, "MapReduce : Simplified Data Processing on Large Clusters," vol. 51, no. 1, pp. 107–113, 2008.
36. A. Gersho, "Vector Quantization : A Pattern-Matching Technique for Speech Coding," no. December, 1983.
37. 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.
38. J. E. Dalton and W. A. Benish, "An Information-Theoretic Measure for Balance Assessment in Comparative Clinical Studies," 2020.
39. J. Zhong, R. Liu, and P. Chen, "Identifying critical state of complex diseases by single-sample Kullback – Leibler divergence," pp. 1–15, 2020.
40. E. Kartal and K. Hamparsum, "Model selection in multivariate adaptive regression splines (MARS) using information complexity as the fitness function," *Mach. Learn.*, pp. 35–58, 2015.
41. H. Bozdogan and J. A. Howe, "Misspecified Multivariate Regression Models Using the Genetic Algorithm and Information Complexity as the Fitness Function," vol. 5, no. 2, pp. 211–249, 2012.
42. J. Canul-Reich, L. O. Hall, D. B. Goldgof, J. N. Korecki, and S. Eschrich, "Iterative feature perturbation as a gene selector for microarray data," *Int. J. Pattern Recognit. Artif. Intell.*, vol. 26, no. 5, pp. 1–25, 2012.
43. I. S. D. and G. B. Carpentieri1, "Sparse pattern selection strategies for robust Frobenius-norm minimization preconditioners in electromagnetism," no. December 1999, pp. 667–685, 2000.
44. T. Hastie and J. D. Lee, "Matrix Completion and Low-Rank SVD via Fast Alternating Least Squares," vol. 16, pp. 3367–3402, 2015.
45. E. Pamukcu and E. Pamukcu, "Choosing the optimal hybrid covariance estimators in adaptive elastic net regression models using information complexity complexity," vol. 9655, 2019.
46. D. X. Zhou, "On grouping effect of elastic net," *Stat. Probab. Lett.*, vol. 83, no. 9, pp. 2108–2112, 2013.
47. T. R. Golub *et al.*, "Molecular Classification of Cancer : Class Discovery and Class Prediction by Gene Expression Monitoring," vol. 286, no. October, pp. 531–538, 1999.
48. D. Singh *et al.*, "Gene expression correlates of clinical prostate cancer behavior," vol. 1, no. March, pp. 203–209, 2002.
49. Z. Y. Algamal and M. H. Lee, "Regularized logistic regression with adjusted adaptive elastic net for gene selection in high dimensional cancer classification," *Comput. Biol. Med.*, vol. 67, pp. 136–145, 2015.