

Classification of high-dimensional data using Penalized Regression Methods under diverse correlation structures: A simulation and real-data based study

Shambhavi Mishra^a, Sakshi Gupta^{b*}, Madhulika Dube^c

*Corresponding author: Sakshi Gupta

Research Scholar, Department of Statistics, University of Lucknow-226007

Email.Id: sakshigstat@gmail.com

Shambhavi Mishra: Assistant Professor, Department of Biostatistics & Health Informatics, SGPGIMS

ORCID ID: <https://orcid.org/0000-0002-7057-067X>

Sakshi Gupta: Research Scholar, Department of Statistics, University of Lucknow

ORCID ID: <https://orcid.org/0009-0000-2082-1939>

Madhulika Dube: Professor, Department of Statistics, BBAU

Abstract

Solving the problem of classification in high-dimensional situations presents substantial challenges because of the proliferation of variables, many of which may be irrelevant or redundant, thereby increasing the risk of overfitting and diminishing predictive performance. As multicollinearity, overfitting, and bias pose serious problems for conventional classification algorithms, variable selection strategies are significant while dealing with high-dimensional data. To address the issue of classification in high-dimensional data under diversified multicollinearity structured data four penalized methods- LASSO, Group LASSO, Adaptive LASSO, and Elastic net have been studied in this paper. Analyses of considered methods are performed on both simulated data with varying correlation structure and real-life data (Prostate cancer data (singh, 2002)), and performance measures such as F1 score, misclassification rate, and variable selection are examined. In simulation, keeping the sample size and within-group correlation constant under varying correlation structures of the data the performance of considered methods was assessed with dimensionality increase. The results revealed that LASSO balanced the sparsity and complexity while maintaining the generalizability of the model. LASSO showed a high F1 score in simulated data, and accuracy increased with multicollinearity in the data. Adaptive LASSO improved variable selection and turned out as most parsimonious model. In simulation studies, Elastic net has shown consistent model selection in penalized regression methods with the highest F1 score and low misclassification rate. In gene expression data, Elastic net consistently selected the best model by achieving the highest F1 score and minimum misclassification error rate but failed to maintain sparsity in both conditions.

Keywords: Penalized logistic regression, high-dimensional data, simulation, regularization, k-fold cross-validation.

How to cite this article: Shambhavi Mishra, Sakshi Gupta*, Madhulika Dube. Classification of high-dimensional data using Penalized Regression Methods under diverse correlation structures: A simulation and real-data based study. *Int J Drug Deliv Technol.* 2026;16(83s):1648-1658. DOI: 10.25258/ijddt.16.83s.180.

Introduction

Classification is an important problem in supervised machine learning and statistical modeling, with applications spanning various domains such as finance, healthcare, image recognition and natural language processing. However, achieving high precision in classification while maintaining model generalization remains a critical challenge (Hastie et al. 2009). The high-speed digitalization of society and advancements in measurement and storage technologies have resulted in an overwhelming influx of data (Chang et al. 2014). The use of high-end computing in the medical field is increasing rapidly every day due to method advancements, demanding the development of a highly accurate classifier system (Du 2014). An additional issue with high dimensionality is that there are more genes k , than observations n (samples) (Cui et al. 2013). Working with the problem $k \gg n$, known as high-dimensional data, creates one more complicated task in the traditional classification methods (Piao 2012).

The most common problem impacting classification models when working on high-dimensional data is overfitting, which happens when a model performs good on training data but is unable to generalize to new data. The problem of overfitting and multicollinearity occurs in high-dimensional datasets, making the classification process more difficult (Chen 2014).

Cancer classification is one such problem where, biologically, only a subset of genes contributes to the identification of cancer disease, and most of the genes remain irrelevant. These extraneous genes add noise while decreasing the precision of the classification (Chandra and Gupta 2011). To overcome such types of problems, effective variable selection methods are needed for the classification process, which reduces the dimension of data and also improves the prediction accuracy. Identification of relevant genes (variables) is critically important in early cancer diagnosis and drug discovery, as

these genes serve as the biomarkers for disease progression (Chen 2014). In machine learning and statistical modeling, variable selection is a critical stage, especially when working with high-dimensional data. The existence of irrelevant or redundant characteristics can cause overfitting, increased computing complexity, and poor interpretability. Several statistical methods have been used successfully to classify cancer and other high-dimensional data sets. Among these, by far, penalized regression methods are considered powerful classification methods. Penalized logistic regression (PLR) performed well in high-dimensional data of cancer classification to address both gene selection and gene coefficient estimation at the same time (Zhu 2004). The l_1 -penalty, also stated as the “least absolute shrinkage and selection operator (LASSO)”, is the most popular and commonly used regularization method (Tibshirani 1996). The LASSO applies the l_1 -penalty to the loss function and can perform variable selection by setting some coefficients to zero due to l_1 -penalty property. As a result, the LASSO is commonly utilized with high-dimensional data. A sparse model with good classification accuracy can be produced by PLR with l_1 -penalty (Algamal and Lee 2015). Using different penalties in PLR can also produce more accurate and significant results. To study the penalization structure of penalization methods on various data conditions, four regularization methods- LASSO, Group LASSO, Adaptive LASSO, and Elastic net are considered for the experiment.

In this paper, PLR methods weighted by different penalties are considered to study the effect of regularization on the high-dimensional data. The other sections of the paper are outlined as follows: Several papers pertaining to regularization techniques are listed in Section 2, while the methods used in the study are delineated in Section 3. In Section 4, a simulation study is conducted to compare the considered methods, and comparison of these methods on the real dataset is presented in Section 5, along with an introduction of the real-life dataset, with the discussion of the results in Section 6. At the end, in Section 7, the main conclusion is drawn. All computations were performed using R software on version 4.2.3.

1. Review of Literature

Regularization is a dimension-reduction method in machine learning that incorporates a regularization term to the loss function to prevent overfitting. By discouraging exceptionally complex models, they improve model stability, reduce overfitting, and enhance predictive performance, leading to better generalization on unseen data (Ng A. Y. 2004). Among the different approaches to the selection of variables pin high-dimensional data, regularization techniques have been employed extensively [Yuan

and Lin (2006), Zou and Hastie (2005), Tibshirani (1996)].

Among several methods introduced to date, Ridge regression was the first, in which the l_2 norm used as the penalty term, which shrinks the coefficients in the direction of zero and reduces the effect of multicollinearity among explanatory variables. But the l_2 norm not set the value of coefficients to exactly zero; therefore, this method fails to decrease the dimensions of the data. Due to the drawback of the l_2 norm, a new technique was introduced known as “LASSO (Least Absolute Shrinkage and Selection Operator)”, which integrates l_1 norm to the loss function (Tibshirani 1996). Although LASSO method has many advantages, this method has some limitations (see Wang et al. 2011; Zheng and Liu 2011). First, it is constrained by its inability to select the variables more than n , and second, when variables show the grouping patterns have high pairwise correlations, in such cases, LASSO typically selects only one or a small subset of genes from a highly correlated group, potentially overlooking biologically relevant features (Algamal and Lee 2015). Group LASSO (GLASSO) regression is an improved version of the LASSO regression method, designed to handle the grouping structure of the variables by selecting or shrinking entire groups of predictors together rather than individual coefficients (Yuan & Lin 2006). GLASSO imposes penalty for selecting the variables at the group level, which means that it either includes or excludes entire groups of variables, restricting flexibility when only a certain number of variables within a group are significant (Huang et al. 2009). To overcome the last limitation of LASSO, Zou (2006) introduced the adaptive LASSO technique in which the initial LASSO estimators are used as adaptive weights for applying penalties on different coefficients. To overcome the drawbacks of LASSO, another technique Elastic net was proposed, in which a combined weight of the l_1 -norm and l_2 -norm is used for penalizing the coefficients (Zou and Hastie 2005).

2. Methodology

The methodology used in this paper is sequenced as, first, the data simulation has been done, and the performances of four considered penalized methods are compared on the simulated data. And again, the evaluation of classification accuracy of the considered methods is tested on the real-life data (prostate cancer data, singh 2002). The methodological framework consists of data generation, model implementation, parameter tuning, and performance evaluation.

3.1. Penalized Logistic Regression Methods

Penalized LR is an extension of logistic regression that incorporates a non-negative regularization term in the loss function, allowing for control over the magnitude of regression coefficients in classification. In penalized regression methods,

generally l_1 and l_2 norms are used as the penalty terms. The norm-based penalty terms are integrated with the loss function to perform shrinkage and selection methods. Let $y \in \{0,1\}$ be the $n \times 1$ response vector, and x_j be the vector of explanatory variables. Logistic transformation of vector of estimates of probability $p_i = P(y_i = 1|x_j)$ is modeled by linear function. The log-likelihood of Logistic regression is given by,

$$l(\beta_0, \beta) = - \sum_{i=1}^n [y_i \ln(p_i) + (1 - y_i) \ln(1 - p_i)] \quad (1)$$

The probability estimates of $P(y_i = 1|x_j)$ is given by,

$$\hat{p}_i = P(y_i = 1|x_j) = \frac{e^{\beta_0 + \sum_{j=1}^k x_{ij}^T \beta_j}}{1 + e^{\beta_0 + \sum_{j=1}^k x_{ij}^T \beta_j}} \quad (2)$$

Equation (2) helps in constructing the predicted class for the variables that have $\hat{p}_i > 0.5$.

3.1.1. Ridge Logistic Regression

Ridge regression is attained by applying penalized parameter l_2 norm, also known as Euclidean norm, to all the coefficients except the intercept (Hoerl 1970). The negative log-likelihood function of Ridge logistic regression is given by

$$l_\lambda^R(\beta) = - \sum_{i=1}^n [y_i \ln(p_i) + (1 - y_i) \ln(1 - p_i)] + \lambda \sum_{j=1}^k \beta_j^2 \quad (3)$$

The estimates of the coefficients are obtained by minimizing the cost function:

$$\hat{\beta}_R = \arg \min_{\beta} [- \sum_{i=1}^n \{y_i \ln(p_i) + (1 - y_i) \ln(1 - p_i)\} + \lambda \sum_{j=1}^k \beta_j^2] \quad (4)$$

where, λ be the tuning parameter and has to be calculated independently, which balances between bias and variance. To reduce the misclassification error of the considered methods, the tuning parameter has to achieve an optimal balance between bias and variance within the range (Hastie et al. 2009).

3.1.2. LASSO Logistic Regression

LASSO is an important method to deal with classification problems, particularly in high-dimensional settings. By incorporating an l_1 -norm penalty into the objective function, the LASSO methods achieve both coefficient shrinkage and variable selection. The objective function to be maximized (or equivalently, the loss function to be minimized) is:

$$\arg \min_{\beta} [- \sum_{i=1}^n \{y_i \ln(p_i) + (1 - y_i) \ln(1 - p_i)\} + \lambda \sum_{j=1}^k |\beta_j|] \quad (5)$$

Due to its computational efficiency, the LASSO method offers a significant advantage in high-dimensional classification problems (Zou, 2006).

3.1.3. Group LASSO Logistic Regression

GLASSO is the modified version of LASSO regression designed to handle grouped predictors. It applies an l_2 -norm penalty at the group level and an l_1 -norm penalty across groups, allowing selection or exclusion of entire groups of correlated predictors collectively. The convex function for GLASSO Regression is:

$$\arg \min_{\beta} [- \sum_{i=1}^n \{y_i \ln(p_i) + (1 - y_i) \ln(1 - p_i)\} + \lambda \sum_{l=1}^L \sqrt{s_l} \|\beta_l\|_2] \quad (6)$$

Where, $\sqrt{s_l}$ be the weights assigned to the coefficients, which are the group sizes of the correlated variables (Friedman et al. 2010). The grouping of the variables has been done using the hierarchical clustering method.

3.1.4. Adaptive LASSO Logistic Regression

To minimize selection bias and obtain a consistent model, adaptive LASSO (adaLASSO) assigns higher weights to smaller coefficients and lower weights to larger coefficients. The adaLASSO model is,

$$\arg \min_{\beta} [- \sum_{i=1}^n \{y_i \ln(p_i) + (1 - y_i) \ln(1 - p_i)\} + \lambda \sum_{j=1}^k \hat{w}_j |\beta_j|] \quad (7)$$

where, $\hat{w}_j = (|\hat{\beta}_j|)^{-\gamma}$, γ is the positive constant, and $\hat{\beta}_j$ be the initial estimate of β_j obtained through Ridge estimation. Following (Alagmal, 2015), the constant $\gamma=1$ taken in the analysis process.

3.1.5. Elastic net Logistic Regression

Elastic net method is the regularized method that merges both l_1 and l_2 norms to balance stability and sparsity simultaneously. Such a combination effectively handles highly correlated variables while simultaneously performing variable selection.

$$\hat{\beta}_{EN} = \arg \min_{\beta} [- \sum_{i=1}^n \{y_i \ln(p_i) + (1 - y_i) \ln(1 - p_i)\} + \lambda \{(1 - \alpha) \sum_{j=1}^k |\beta_j| + \alpha \sum_{j=1}^k \beta_j^2\}] \quad (8)$$

Elastic net estimates depend on λ and α the non-negative parameters, that leads to regularization.

3.2. Tuning of parameters

The regularization parameter for LASSO, GLASSO, and Adaptive LASSO were fitted using k-fold cross-validation method. The parameters λ and α for Elastic net were tuned using grid search combination in which those α values are considered for the model, which minimized the misclassification error and simultaneously balanced the variable selection. All models were fitted using coordinate descent optimization algorithm.

3.2.1. k-fold Cross-Validation

The validation method to find the best tuning parameter λ selection is k-fold cross-validation. This method divides the set of observations into k groups at random. The first fold works as the validation set, while subsequent k-1 folds taken to fit the model. With this configuration, each fold functions as a validation set once since the process

has to be performed k times. In the classification problem, misclassification errors $Err_1, Err_2, Err_3, \dots, Err_k = I(y_i \neq \hat{y}_i)$. The k -fold CV estimate computed by averaging the values,

$$CV_k = \frac{1}{k} \sum_{i=1}^k Err_i \quad (9)$$

k -fold cross-validation is beneficial in high-dimensional data sets and considered a standard method in penalized regression (James et al. 2021; Hastie et al., 2009).

3.2.2. Coordinate Descent Algorithm

Following Friedman et al. (2010), the coordinate descent algorithm used to compute the regularization path. According to the penalties applied on the loss functions of the regularization methods, the objective functions and the coordinate update are given by

Table 1: The objective functions of the methods and their coordinate update

Method	Objective function	Coordinate update
LASSO	$\frac{1}{2n} \ Y - X\beta\ _2^2 + \lambda \sum_{j=1}^k \beta_j $	$\hat{\beta}_j \leftarrow S\left(\frac{1}{n} X_j^T r_j, \lambda\right)$
Group LASSO	$\frac{1}{2n} \ Y - X\beta\ _2^2 + \lambda \sum_{s=1}^S \ \beta_s\ $	$\hat{\beta}_s \leftarrow \left(1 - \frac{\lambda}{\ z_s\ _2}\right)_+ z_s$
Adaptive LASSO	$\frac{1}{2n} \ Y - X\beta\ _2^2 + \lambda \sum_{j=1}^k w_j \beta_j $	$\hat{\beta}_j \leftarrow S\left(\frac{1}{n} X_{ji}^T r_j, \lambda w_j\right)$
Elastic net	$\frac{1}{2n} \ Y - X\beta\ _2^2 + \lambda \left[\alpha \sum_{j=1}^k \beta_j + (1 - \alpha) \sum_{j=1}^k \beta_j^2 \right]$	$\hat{\beta}_j \leftarrow \frac{S\left(\frac{1}{n} X_j^T r_j, \lambda \alpha\right)}{(1 + \lambda(1 - \alpha))}$

where, $r_j = (y - X_{-j}\beta_{-j})$.

The coordinate descent algorithm maintains all other coefficients constant while minimizing the objective function in respect to a single coefficient, β_j .

3. Simulation

In the paper, the simulation experiment was performed to assess the performance of different regularization techniques in high-dimensional classification problems. The data was simulated using the logistic transformation of the probability

vector estimates $\pi_i = p(y_i = 1|x_i)$, and represented by a linear predictor with the logit link function. The logit transformation is given by:

$$\ln \left[\frac{\pi_i}{1 - \pi_i} \right] = \beta_0 + \sum_{j=1}^k x_{ij}^T \beta_j, \quad i = 1, 2, \dots, n$$

Logistic regression has the ability to estimating both the probability p_i and $(1 - p_i)$ for each class and categorizing observations based on them. The probability of assigning the i^{th} sample to class 1 is given by

$$\hat{p}_i = P(y_i = 1|x_i) = \frac{1}{1 + e^{-(\beta_0 + \sum_{j=1}^p x_{ij}^T \beta_j)}}$$

The decision rule $I\{\hat{p}_i > 0.5\}$ was used to predict the class label. The indicator function $I(\cdot)$ gives 1 to a favorable condition and 0 otherwise.

4.1. Simulation design

In evaluation of the regularized logistic regression methods, concentrating on the high-dimensional settings and varying correlation structures, three different simulation designs are conducted. The binary response variable y_i was produced from a Binomial distribution with probability p_i , and the predictor matrix $X \in R^{n \times k}$ was generated from the normal distribution. To study the effect of multicollinearity, five different values of the correlation parameter $\rho = \{0.0/0.2/0.4/0.8\}$ were considered in each simulation design. Data generated for every simulation design by keeping the sample size ($n=100$) and within-group correlation r constant and for result validation training and testing has been done by dividing the data into two parts (70% train data, 30% test data). All the simulation designs replicated (m) 100 times to obtain the consistent results.

Simulation design 1: In the first simulation setting, 1000 predictors were taken, with within-group correlation $r = 0.3$. The true beta coefficients for the first design were

$$\beta_{k \times 1} = \{\underbrace{1, 1, \dots, 1}_{100}, 0, 0, \dots, 0, 0\}$$

Only 100 beta coefficients were kept non-zero by assigning the value 1 and remaining set to zero.

Simulation design 2: In the second simulation setting, 2000 predictors were taken, with within-group correlation $r = 0.3$. The true beta coefficients for the first design were

$$\beta_{k \times 1} = \{\underbrace{1, 1, \dots, 1}_{100}, 0, 0, \dots, 0, 0\}$$

Only 100 beta coefficients were kept non-zero by assigning the value 1 and remaining set to zero.

Simulation design 3: In the third simulation setting, 3000 predictors were taken, with within-group correlation $r = 0.3$. The true beta coefficients for the first design were

$$\beta_{k \times 1} = \{\underbrace{1, 1, \dots, 1}_{100}, 0, 0, \dots, 0, 0\}$$

Only first 100 beta coefficients were kept non-zero by assigning the value 1 and remaining set to zero.

4.2. Simulation Results

The influence of dimensionality of data on all four PLR methods with different levels of multicollinearity has been evaluated on simulated data. By maintaining the sparsity (true non-zero coefficients = 100) and constant low within-group correlation ($r = 0.3$) in the data, classification performances of the methods have been assessed using the F1 score and misclassification error (Misc_err). The average of the F1 score and misclassification error on 100 replications has been taken for comparison. The findings show that dimensionality and correlation structure (ρ) both have a significant impact on the performances of variable selection methods.

4.2.1. Comparative assessment of penalized methods by predictive performance

The predictive performances of all four methods are presented in fig. 1. and the performances are compared by F1 score and misclassification error.

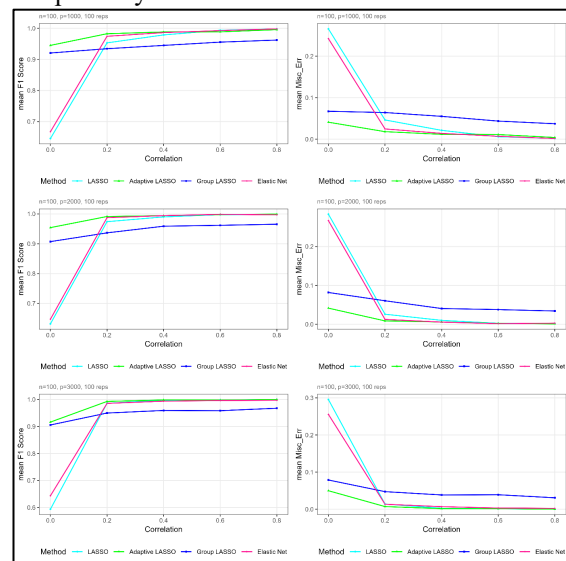


Fig. 1. Classification accuracy average by F1 score and misclassification error with different levels of correlation

The results of Table 1. indicate that when the data has only within-group correlation, LASSO and Elastic Net both struggle in correctly identifying the relevant variables, and also, misclassification error (MCE) remains high as compared to the other two methods. For $\rho = 0.0$, LASSO and Elastic Net attain low F1 scores with values 0.6455 and 0.6673, respectively. Despite having nearly the same characteristics, adaLASSO performed better in identifying the true positive cases and attained a minimum MCE (0.0409) among all the methods. As in the first case, only the group correlation exists, so GLASSO appropriately carried out the within-group correlation structure of the data, suggesting that grouping information contributes positively to stable variable selection. Conventional LASSO and Elastic Net performed considerably worse in the absence of

correlation but dramatically improved the classification performance for $\rho = 0.2$ by decreasing the MCE from 0.2657 to 0.0463 for LASSO and from 0.2421 to 0.0247 for Elastic Net. This substantial improvement suggests that moderate correlation among predictors facilitates the identification of the true sparse structure by stabilizing coefficient estimation and enhancing shrinkage efficiency. AdaLASSO consistently performed well by maintaining a high F1 score and low MCE in all the considered correlation structures. GLASSO remained stable with severity in the multicollinearity; it indicates that GLASSO is less sensitive to the level of multicollinearity. Although GLASSO demonstrated a relatively slower improvement. As multicollinearity among the independent variables increased, the consistency remained strong by maintaining the high F1 score, yet MCE did not decrease noticeably as for $\rho = 0.0$ it was 0.0671, and for $\rho = 0.8$ it decreased only by 0.03, which was the lowest among all the methods. LASSO and Elastic Net showed comparatively similar behavior over the degree of multicollinearity and also increased the dimensionality. As the evidence displays, these penalized methods have a strong dependence on the correlation structure and dimensionality of the data. Declining the MCE with increment in multicollinearity depicts that the shrinkage and selection mechanism of the LASSO works more effectively in the selection of relevant variables and reducing estimation variability. However, the l_1 penalty shrinks coefficients aggressively and may incorrectly remove the important predictors along with the irrelevant variables. Therefore, the model built in presence of strong collinearity between the predictor variables is suitable for predicting the outcome based on the predictor variables. Nevertheless, it is not useful for drawing conclusions. about how each predictor variable and the response variable are related. Whereas Elastic Net shares the combination of l_1 and l_2 both penalties, therefore, by reducing the smaller coefficients towards zero and selecting the relevant variables, it encouraged the grouped selection of correlated variables and stabilized the estimation with the increasing severity of multicollinearity. As in adaLASSO initial weights are taken by the ridge estimation, which reduced the irrelevant variable coefficients towards zero; after that, adaLASSO was performed. AdaLASSO gave consistently satisfactory results under the independent predictors setting ($\rho = 0.0$) with F1 score = 0.9452 and also in the high multicollinearity conditions ($\rho = 0.8$) with F1 score = 0.9958, it also maintained low MCE over all correlation settings. It demonstrated the robustness and stability in high-dimensional sparse regression problems.

Table 2. Classification performance results on different simulation designs

Method	$\rho = 0.0$		$\rho = 0.2$		$\rho = 0.4$		$\rho = 0.6$		$\rho = 0.8$	
	F1 score	Misc_err	F1 score	Misc_err	F1 score	Misc_err	F1 score	Misc_err	F1 score	Misc_err
For $p = 1000, n = 100, r = 0.3$										
LASSO	0.6455	0.2657	0.9532	0.0463	0.9788	0.0213	0.9938	0.006	0.9983	0.0017
Group LASSO	0.9207	0.0671	0.9345	0.0642	0.9451	0.055	0.9556	0.0436	0.9622	0.0372
AdaLASSO	0.9452	0.0409	0.9827	0.0181	0.9883	0.0117	0.9887	0.0112	0.9958	0.0041
Elastic net	0.6673	0.2421	0.9744	0.0247	0.9863	0.014	0.9927	0.007	0.9983	0.0016
For $p = 2000, n = 100, r = 0.3$										
LASSO	0.6314	0.2832	0.9739	0.0259	0.9899	0.0101	0.9974	0.0025	0.9982	0.0017
Group LASSO	0.907	0.0818	0.9364	0.0605	0.9587	0.0405	0.9617	0.0378	0.9653	0.0342
AdaLASSO	0.954	0.0415	0.9912	0.0087	0.994	0.0059	0.9977	0.0021	0.9989	0.0011
Elastic net	0.647	0.2667	0.9873	0.0125	0.9941	0.0056	0.9985	0.0015	0.9968	0.0029
For $p = 3000, n = 100, r = 0.3$										
LASSO	0.5941	0.2955	0.9857	0.2955	0.9975	0.0027	0.9981	0.0015	0.9996	0.0004
Group LASSO	0.9055	0.0789	0.9497	0.0789	0.9592	0.0386	0.9584	0.0392	0.9675	0.031
AdaLASSO	0.9163	0.05	0.9926	0.05	0.9983	0.0018	0.9978	0.0018	0.9997	0.0003
Elastic net	0.6437	0.2553	0.9853	0.2553	0.9936	0.0071	0.9968	0.0034	0.9969	0.0026

In the context of dimensionality, all the methods except GLASSO are impacted by high-dimensionality. LASSO demonstrated relatively modest performance in a lower-dimensional setting, and the performance got worse with increment in the dimensionality ($p = 3000$, F1 score = 0.5941, Misc_err = 0.2955). This happened because of the nature of LASSO to arbitrarily pick one variable from correlated groups and produce unstable selection under high sparsity. However, the considerable improvement noticed in the performance of LASSO with the increase in multicollinearity. GLASSO utilized the within-group correlation information and merely changed the F1 score for $\rho = 0.0$ from 0.9207 ($p = 1000$) to 0.9055 ($p = 3000$) and maintained the low MCE from 0.0671 ($p = 1000$) to 0.0789 ($p = 3000$). As the correlation increased, these changes did not remain significant, in the high correlation structure, GLASSO performed better with the increment in the dimensionality. AdaLASSO consistently achieves the higher F1-scores across all dimensional settings and the low misclassification errors (Misc_err = 0.0003 at $p = 3000$), outperforming all competing approaches. Because adaLASSO uses a data-driven adaptive weighting approach, which gives each coefficient a varying penalty strength. The simulation results unmasked the redundant approach of the Elastic Net in the absence of multicollinearity, as the dimension of the data increased ($p = 2000$) the MCE also increased (Misc_err = 0.2667) but for $p = 3000$, it remained stable yet high. After introducing the multicollinearity in the simulated data, Elastic net improved its predictive accuracy drastically. It might be due to hybrid l_1 and l_2 penalty structure, which helps Elastic Net in simultaneous shrinkage of irrelevant variables and selection of relevant variables.

4.2.2. Strength of penalization and variable selection

As Fig. 2 represents, adaLASSO imposed the highest penalty on the variables among all the methods and as a result, selected the fewest variables in the multicollinearity setting. Elastic Net used highest penalization in independence yet performed the worst, and the same pattern is followed by LASSO. However, GLASSO nearly equally regularized in all the considered simulation designs

and approximately identified equal non-zero coefficients.

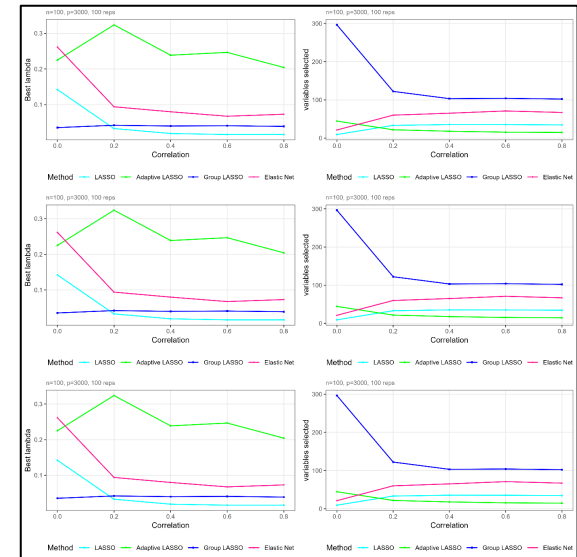


Fig. 2. Penalties imposed and number of variables selected by penalization methods

The outcomes presented in Table 2. highlighted that LASSO posed a higher penalty across all the dimensions from 0.1222 to 0.1421 in the absence of correlation and selected the least number of variables (12, 10, and 9) among all the penalized methods. This is because, in the absence of multicollinearity, l_1 penalty imposes strong shrinkage, destructively eliminating a large proportion of predictors and retaining arbitrarily only one variable from the correlated group. While, with the introduction of multicollinearity in the data, penalty reduced substantially and selected significant non-zero coefficients due to shrinkage mechanism in the sparse high-dimensional framework. GLASSO illustrated stable regularization values (0.03-0.04), which means that it did not over or under penalized the data. But it selected suspiciously a large number of variables (228, 240 and 296), especially in $\rho = 0.0$ setting, and also faced the curse of dimensionality. Since GLASSO uses the group-wise penalization methodology, in which l_2 norm of the group coefficients becomes small, making it difficult for the penalty to shrink entire groups to zero. AdaLASSO consistently displayed the most precise and parsimonious variable selection behavior across all experimental conditions. Although it used high lambda values consistently in producing sparse models. The monotonic decline in the number of variables chosen as ρ increases is a noteworthy trend. AdaLASSO chose 40 variables at $\rho = 0.0$ at $p=1000$ but reduced this to only 12 variables at $\rho=0.8$. This happened due to the ridge penalty which used as the initial weights, and the ridge penalty encourages correlated active predictors to be jointly retained, leading to a rise in the dimension of the variables and as the overall correlation increases,

adaLASSO imposes the l_2 penalty and selects fewer variables from the groups.

Table. 3. Penalties imposed and number of variables selected under different simulation design

Method	$\rho = 0.0$		$\rho = 0.2$		$\rho = 0.4$		$\rho = 0.6$		$\rho = 0.8$	
	best.lambd a	Var. selecte d	best.lambd a	Var. selecte d	best.lambd a	Var. selecte d	best.lambd a	Var. selecte d	best.lambd a	Var. selecte d
For $p = 1000, n = 100, r = 0.3$										
LASSO	0.1222	12	0.0509	24	0.0352	26	0.0234	29	0.0145	32
Group LASSO	0.0333	228	0.0414	109	0.0417	82	0.0403	80	0.0378	86
AdaLASSO	0.3924	40	0.4237	23	0.3808	17	0.3967	13	0.2665	12
Elastic net	0.2242	22	0.0929	50	0.076	54	0.0633	58	0.0614	57
For $p = 2000, n = 100, r = 0.3$										
LASSO	0.1365	10	0.0441	28	0.0273	31	0.0182	34	0.014	33
Group LASSO	0.0377	240	0.0433	104	0.0399	95	0.0409	94	0.0379	99
AdaLASSO	0.2805	43	0.337	23	0.2848	18	0.2534	15	0.2287	13
Elastic net	0.2584	18	0.085	62	0.0726	65	0.0653	65	0.0672	64
For $p = 3000, n = 100, r = 0.3$										
LASSO	0.1421	9	0.0329	33	0.019	35	0.0162	35	0.0162	34
Group LASSO	0.0356	296	0.0421	122	0.04	103	0.0408	104	0.0389	102
AdaLASSO	0.225	45	0.3237	22	0.2387	18	0.2466	15	0.2042	15
Elastic net	0.2615	21	0.0939	60	0.08	65	0.0675	71	0.0731	67

5. Application

5.1. Real-Life Dataset

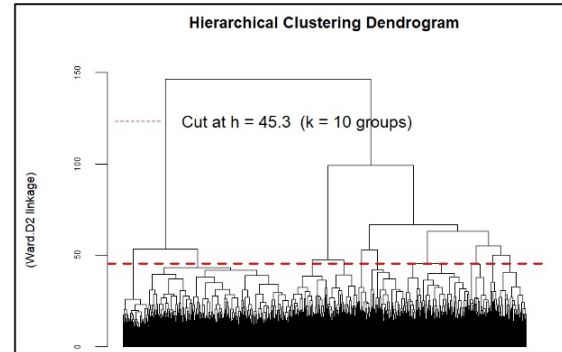
A sample of 6033 genes for 50 non-tumor patients and 52 tumor patients was taken from the 12,600 genes' expression of the Prostate Cancer dataset (singh et al., 2002) from the R repository. The data have two elements, in which "y" is the classification of the sample (normal, tumor) and "X" is the matrix with 102 observations and 6033 gene expressions. Again, this data was split into a 70:30 proportion, and the penalized methods considered under study were first applied to train the model, and predictions were made on test data.

5.2. Results

The results of all four methods are presented in Table. 3. Since the data was high-dimensional and imbalanced therefore, F1 score and misclassification errors are used to compare the models' performance. Also, the penalty imposed by different methods and sparsity maintained by the models are compared.

LASSO: Efficient Sparsity and variable selection
LASSO is treated as the baseline model for comparing the performances of the considered methods on real data. As mentioned in the name, LASSO maintained sparsity and balanced the penalty to perform the variable selection. The model returned an F1 score of 0.7741 and an MCE of 0.2258, establishing a solid foundation for feature selection. The algorithm effectively compressed the dimensionality of the data by imposing the penalty $\lambda = 0.0263$ and selected 43 genes. While LASSO is highly efficient at identifying independent predictors by forcing less relevant coefficients to zero, it lacks in selecting the true non-zero coefficients; therefore, it does not follow the oracle property.

Group LASSO: Structural overfitting and noise retention



Since initially the correlation structure of the data was not known. Hence, to group the variables on the basis of correlation exist between them, hierarchical clustering with Ward's linkage method was used. On performing hierarchical clustering on gene expressions, a dendrogram formed presented in Fig. 4., gave the cut at $k = 10$ groups. Therefore, 10 groups of correlated gene expressions were obtained.

Fig. 3. Hierarchical clustering Dendrogram of gene expression dataset

After clustering the variables, GLASSO was performed on 10 groups of variables, and it yielded the least effective performance, with an F1 score of 0.6667 and a high MCE of 0.4194. The GLASSO selected a great number of variables (2942) and lost sparsity of the model. Since GLASSO inherently selects the whole group of variables on the basis of a few relevant variables, irrespective of the degree of relevance. The model introduced excessive noise into the decision boundary by keeping whole blocks of variables regardless of their individual relevance. This resulted in severe overfitting and a poor ability to generalize to the classification.

Adaptive LASSO: Most parsimonious model

By utilizing the data-driven LASSO weights in the initial phase, adaLASSO produced the sparsest model by selecting the fewest variables (30). Since on taking ridge estimators as the initial weights, the lambda value taken by adaLASSO was very high and overfitted the data; therefore, LASSO estimators are used as the initial estimates, although it still imposed the highest penalty ($\lambda = 0.0921$) among all the methods. The performance measure F1 score was 0.75, which was near to LASSO, as it addresses the bias issues inherent in LASSO, but also satisfies the oracle property, therefore selected the true non-zero coefficients.

Elastic net: Highest prediction accuracy and stability

With an MCE of 0.1290 and the highest F1 score of 0.8667, the Elastic Net approach proved to be the most reliable model in this experiment. These outcomes illustrate the efficacy of its hybrid regularization strategy. Elastic Net works well in genomic research, where genes frequently show high multicollinearity because of similar biological pathways. With 62 genes chosen, the model strikes a smart compromise between being inclusive

enough to capture the intricate, synergistic signals required for high-accuracy classification and being sparse enough to be comprehensible.

Table 3. Classification results of Real-life data:

Method	F1	MCE	Lambda	No. of
LASSO	0.774	0.225	0.0263	43
Group	0.666	0.419	0.0562	2942
AdaLASS	0.75	0.258	0.0921	30
ELASTIC	0.866	0.129	0.0576	62

6. Discussion

The comparative assessment of LASSO, GLASSO, adaptive LASSO, and Elastic net was conducted on varying correlation-structured simulated designs and gene expression data to assess their classification performance, misclassification error rates, and variable selection capabilities. Under various data-driven conditions, the experimental findings disclose significant variations in variable selection consistency, predictive accuracy, and sparsity maintenance among different approaches, with the performance outcomes significantly influenced by the structural characteristics of the predictor space. The results demonstrated that LASSO performed well in terms of sparsity and complexity simultaneously as the variables increased. However, a major drawback of LASSO is that, due to convex optimization nature, it can choose a maximum of n variables when $k > n$. Second, when predictors show high grouped pairwise correlation, LASSO faces difficulties with the grouping effect, which means that it typically picks only one variable from the group without taking into account which particular variable is chosen (Zou & Hastie, 2005). Finally, LASSO do not possess oracle properties, as established by (Fan & Li, 2001), meaning it do not consistently identify the true underlying model in finite samples. The adaptive penalty structure improves coefficient discrimination while also lowering the probability of major predictor over shrinkage. As a result, AdaLASSO achieved good selection consistency and estimation accuracy regardless of whether the predictors are weakly or strongly associated. By leveraging data-driven adaptive weights derived from ridge regression estimates as initialization in the simulated data and LASSO estimates in gene expression data, adaLASSO imposes differentiated penalty strengths on individual coefficients, assigning stronger shrinkage to coefficients near zero and smaller penalties to those associated with true signals. This mechanism effectively mitigates the uniform over-shrinkage problem of LASSO, enabling adaLASSO to satisfy the oracle property under appropriate regularity conditions. GLASSO demonstrated stability and sparsity in prior settings (simulated data) but reflected poor generalization in real-life scenario. In high-dimensional gene expression data, GLASSO was preceded by hierarchical clustering with Ward's linkage to

construct ten gene groups. Despite this preprocessing step, the resulting model selected an excessively large number of variables and yielded the lowest F1 score and the highest misclassification error among all methods evaluated. It indicates that the estimates of GLASSO heavily depends on the prior definition of groups; if the groupings are poorly specified or irrelevant to the underlying data structure, the method can select redundant or non-informative variables, leading to suboptimal predictive performance and potential overfitting (Yuan and Lin, 2006). Elastic Net consistently performed regularization very efficiently in both conditions under simulation and real-life dataset (Zou and Hastie, 2005). But a major challenge in Elastic Net is the selection of the value of α for which the misclassification error is minimized; also, it is computationally complex. When multicollinearity is present, Elastic Net provides a better bias-variance trade-off than regular LASSO by simultaneously encouraging sparsity and the grouped selection of correlated predictors. Furthermore, Zou and Zhang (2009) noted that while the grouping effect issue for elastic net exists, it fails to achieve the oracle property. Elastic Net achieved a more delicate response to increasing dimensionality. While its misclassification error remained stable with the change in dimensionality, the introduction of multicollinearity substantially improved predictive accuracy across all dimensional settings. This behavior is consistent with the stabilizing role of the l_2 component in the Elastic Net penalty structure, which facilitates grouped selection and reduces estimation variability as dimensionality grows.

7. Conclusion

The study presented a rigorous and systematic comparative assessment of regularized classification methods- LASSO, GLASSO, adaLASSO, and Elastic net on predefined condition simulated datasets and also real-life data of gene expressions. The F1 score and misclassification error were employed as primary performance metrics to account for the challenges posed by sparse and imbalanced data structures. The outcome of the experiment demonstrate that no single regularized method achieves universal superiority across all experimental conditions. Although Elastic Net turned out to be the best regularization method among those considered in high-dimensional data settings, it struggles in the selection process of variables when the grouping effect exists; also, it is computationally expensive. Throughout all simulation situations and the real-data application, adaLASSO consistently proved to be a trustworthy and adaptable approach. Because of its data-driven weighing technique, it simultaneously provides superior sparsity, consistency in selecting the variables, and prediction accuracy while satisfying oracle property. LASSO balances the sparsity and

complexity while maintaining the generalizability of the model. Yet, it fails to achieve better predictive performance when the variables exhibit the grouping structure. Although it is computationally efficient and interpretable, exhibits inherent instability under correlated and high-dimensional conditions, fails to satisfy the oracle property, and may yield unreliable inferences in settings where the primary goal extends beyond prediction. GLASSO performed well under simulation but showed poor generalization to real-world data, emphasizing its sensitivity to group specification. Moreover, it is highly susceptible to excessive variable inclusion and severe overfitting when groups are derived empirically or when within-group heterogeneity in effect sizes is considerable. Elastic Net consistently delivered strong performance in both simulated and real data sets, effectively managing multicollinearity and selecting informative features. In respect to future research scope, the future research could focus on the development of such regularization methods that can balance the regularization and selection of variables without the prior information of the variables in a high-dimensional dataset while optimizing the computational efficiency. The study reveals that when the primary objective is predictive accuracy and high dimensionality or multicollinearity exists in the data, then penalization methods work as a strong tool for handling multicollinearity and variable selection. While, from the inferential perspective, these methods should be implemented very carefully. Especially, GLASSO should be reserved for applications in which well-defined and externally validated group structures exist a priori, and its use in exploratory settings warrants cautious and critical interpretation. This study can be extended to propose a novel method that can maintain the balance between the prediction accuracy and sparsity in high-dimensional data, simultaneously can draw conclusive inferences.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

References

- Hastie, T., Tibshirani, R., & Friedman, J. H. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed). Springer.
- Chang, R. M., Kauffman, R. J., & Kwon, Y. (2014). Understanding the paradigm shift to computational social science in the presence of big data. *Decision Support Systems*, 63, 67–80. <https://doi.org/10.1016/j.dss.2013.08.008>
- Du, D., Li, K., Li, X., & Fei, M. (2014). A novel forward gene selection algorithm for microarray data. *Neurocomputing*, 133, 446–458. <https://doi.org/10.1016/j.neucom.2013.12.012>
- Cui, Y., Zheng, C.-H., Yang, J., & Sha, W. (2013). Sparse maximum margin discriminant analysis for feature extraction and gene selection on gene expression data. *Computers in Biology and Medicine*, 43(7), 933–941. <https://doi.org/10.1016/j.compbiomed.2013.04.018>
- Piao, Y., Piao, M., Park, K., & Ryu, K. H. (2012). An ensemble correlation-based gene selection algorithm for cancer classification with gene expression data. *Bioinformatics*, 28(24), 3306–3315. <https://doi.org/10.1093/bioinformatics/bts602>
- Chen, K.-H., Wang, K.-J., Wang, K.-M., & Angelia, M.-A. (2014). Applying particle swarm optimization-based decision tree classifier for cancer classification on gene expression data. *Applied Soft Computing*, 24, 773–780. <https://doi.org/10.1016/j.asoc.2014.08.032>
- Chandra, B., & Gupta, M. (2011). An efficient statistical feature selection approach for classification of gene expression data. *Journal of Biomedical Informatics*, 44(4), 529–535. <https://doi.org/10.1016/j.jbi.2011.01.001>
- Zhu, J. (2004). Classification of gene microarrays by penalized logistic regression. *Biostatistics*, 5(3), 427–443. <https://doi.org/10.1093/biostatistics/kxg046>
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 58(1), 267–288. <https://doi.org/10.1111/j.2517-6161.1996.tb02080.x>
- Algarni, Z. Y., & Lee, M. H. (2015). Penalized logistic regression with the adaptive LASSO for gene selection in high-dimensional cancer classification. *Expert Systems with Applications*, 42(23), 9326–9332. <https://doi.org/10.1016/j.eswa.2015.08.016>
- Ng, A. Y. (2004). Feature selection, L1 vs. L2 regularization, and rotational invariance. Twenty-First International Conference on Machine Learning - ICML '04, 78. <https://doi.org/10.1145/1015330.1015435>
- Yuan, M., & Lin, Y. (2006). Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 68(1), 49–67. <https://doi.org/10.1111/j.1467-9868.2005.00532.x>
- Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 67(2), 301–320. <https://doi.org/10.1111/j.1467-9868.2005.00503.x>
- Wang, S., Nan, B., Rosset, S., & Zhu, J. (2011). Random lasso. *The Annals of Applied Statistics*, 5(1). <https://doi.org/10.1214/10-AOAS377>
- Zheng, S., & Liu, W. (2011). An experimental comparison of gene selection by Lasso and Dantzig selector for cancer classification. *Computers in*

- Biology and Medicine, 41(11), 1033–1040.
<https://doi.org/10.1016/j.compbimed.2011.08.011>
16. Huang, J., Ma, S., & Zhang, C.-H. (2008). Adaptive lasso for sparse high-dimensional regression models. *Statistica Sinica*, 18(4), 1603–1618.
<http://www.jstor.org/stable/24308572>
 17. Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American Statistical Association*, 101(476), 1418–1429.
<https://doi.org/10.1198/016214506000000735>
 18. Singh, D., Febbo, P. G., Ross, K., Jackson, D. G., Manola, J., Ladd, C., Tamayo, P., Renshaw, A. A., D’Amico, A. V., Richie, J. P., Lander, E. S., Loda, M., Kantoff, P. W., Golub, T. R., & Sellers, W. R. (2002). Gene expression correlates of clinical prostate cancer behavior. *Cancer Cell*, 1(2), 203–209. [https://doi.org/10.1016/S1535-6108\(02\)00030-2](https://doi.org/10.1016/S1535-6108(02)00030-2)
 19. Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1), 55–67.
<https://doi.org/10.1080/00401706.1970.10488634>
 20. Friedman, J., Hastie, T., & Tibshirani, R. (2010). A note on the group lasso and a sparse group lasso (arXiv:1001.0736). arXiv.
<https://doi.org/10.48550/arXiv.1001.0736>
 21. James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). *An introduction to statistical learning: With applications in r*. Springer US.
<https://doi.org/10.1007/978-1-0716-1418-1>
 22. Fan, J., & Li, R. (2006). *Statistical challenges with high dimensionality: Feature selection in knowledge discovery* (arXiv:math/0602133). arXiv.
<https://doi.org/10.48550/arXiv.math/0602133>
 23. Zou, H., & Zhang, H. H. (2009). On the adaptive elastic-net with a diverging number of parameters. *The Annals of Statistics*, 37(4).
<https://doi.org/10.1214/08-AOS625>