Journal of Biostatistics and Epidemiology

J Biostat Epidemiol. 2024;10(3): 253-272

Original Article

Impact of the Power of Adaptive Weight on Penalized Logistic Regression: Application to Adipocytic Tumors Classification

Narumol Sudjai[1], Monthira Duangsaphon[2*], Chandhanarat Chandhanayingyong[1*]

[1]Department of Orthopaedic Surgery, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok, Thailand.
[2]Department of Mathematics and Statistics, Faculty of Science and Technology, Thammasat University, Pathum Thani, Thailand.

ARTICLE INFO

Received 23.05.2024
Revised 13.07.2024
Accepted 26.07.2024
Published 09.09.2024

Key words:
High-dimensional sparse data;
Machine-learning;
Multicollinearity;
Penalized logistic regression;
Penalty function;
Power of adaptive weight;
Initial weight

ABSTRACT

Introduction: MRI-based texture features in adipocytic tumors to serve as non-invasive predictive biomarkers that can provide precise outcomes for decision-making. Power of adaptive weight and the initial weight for the adaptive Lasso is one of the important parameters. This study aimed to compare the impact of the initial weight together with the power of adaptive weight for this adaptive Lasso under high-dimensional sparse data with multicollinearity.
Methods: All independent variables in the Monte Carlo simulation were generated using the Toeplitz correlation structure. Performance of the initial weight together with the power of adaptive weight on penalized approaches was evaluated using the mean of the predicted mean squared error (MPMSE) for simulation study and the area under the receiver operator characteristic curve (AUC), precision, recall, F1-score, and the classification accuracy of models for real-data applications.
Results: The simulation study showed that the smallest MPMSE value was obtained from the square root of the adaptive Lasso together with the initial weight using Lasso. Additionally, the results of this approach on the real-data application achieved high performance to distinguish the intramuscular lipomas from well-differentiated liposarcomas: the values of AUC, accuracy, precision, recall, and F1-score for the model based on penalized logistic regression classifier were 0.935, 0.928, 0.919, 0.921, and 0.925 respectively, and 0.946, 0.935, 0.932, 0.934, and 0.930 respectively for the model based on support vector machine classifier. Both the simulation study and the real-data application presented that the square root of the adaptive Lasso together with the initial weight using Lasso was the best option under high-dimensional sparse data with multicollinearity.
Conclusion: Our finding showed that the power of adaptive weight on penalty function and the initial weight can affect certain the classification accuracy of machine-learning model. In practice, if choosing these parameters are appropriate, it produces models that have good performance.

Introduction

Advances in technology have enabled computers to effectively store large quantities of the data. Consequently, we desire tools capable of extracting valuable information from massive data sets. To effective decision-making, precise predictive models are essential.

*.Corresponding Authors: chandhanarat.c@gmail.com & monthira@mathstat.sci.tu.ac.th

Logistic regression models are widely applied in data analytics[1, 2] and machine-learning (ML) strategies.[3, 4] This model aims to explain the relationships between independent and outcome variables and predicts future value of the outcome variable. For example, in case where we desire to study the predictive features for distinguishing intramuscular (IM) lipomas from well-differentiated liposarcomas (WDLSs). IM lipomas and WDLSs are benign adipocytic tumors and adipocytic malignancies, respectively. Identifying IM lipomas that are larger than 5 cm, deep-seated, and symptomatic may be difficult because of their resemblance to WDLSs. One of most useful diagnostic tools for these tumors is magnetic resonance imaging (MRI). Thus, the independent variables of interest are MRI-based texture features (these features in adipocytic tumors to serve as non-invasive predictive biomarkers), while the outcome variable would simply be coded as 1 (WDLSs) or 0 (IM lipomas). The statistical method known as binary logistic regression is the one that is commonly utilized in such research. Regarding binary logistic regression coefficients estimation, maximum likelihood estimation (MLE) was regarded as the classical approach. However, this approach is unsuitable when the data of independent variables are high collinear (also called multicollinearity) and small sample size.[5-7] A critical problem that arises in cancer research is a limited sample size, while a data set with a large number of independent variables relative to the sample size (also called high-dimensional data), which can lead to multicollinearity problem. Additionally, the data set may be most of the entries don't contain significant information, which is called "sparse" data.[8] Therefore, high-dimensional sparse data is a challenge in model construction as it can cause overfitting.[9] Another issue is multicollinearity[10], which can cause the variance of logistic regression coefficient estimators to inflate.[5, 6] Consequently, the MLE utilized for estimating logistic regression coefficients is unstable for building a classification model.[11]

In order to remedy the two above problems, penalized approaches can be employed in the logistic regression. These approaches can reduce the variance and help alleviate model overfitting.[12, 13] Popular penalty methods are ridge regression, Lasso, and elastic net.[14-16] These methods are one of techniques that are widely used in machine-learning study.[14-16] Regarding ridge regression, this penalty penalizes the model by shrinking the coefficients toward zero, which can reduce the variance and solve the multicollinearity problem. For Lasso penalty, this method avoids the disadvantage of ridge regression using the tuning parameter controls the shrinkage of the coefficients, which has the effect of shrinking some coefficient estimates to exactly zero. Consequently, this method can remedy the multicollinearity and help alleviate model overfitting. Elastic net combines the properties of ridge regression and Lasso using two tuning parameters to control the shrinkage of coefficients, which has superior performance to Lasso. Although these methods are effective in solving all above problems, they have some disadvantages. For this reason, several earlier studies devoted their attention to devising an adaptive weight for penalized methods.[17, 18] For example, in 2006, Zou[17] proposed adaptive Lasso, which enjoyed oracle properties and led to stable estimation. In 2009, Zou and Zhang proposed an adaptive elastic net[18]; it has oracle properties and superior performance to elastic net. In 2021, Araveeporn[19] compared the higher-order of the adaptive Lasso and adaptive elastic net methods for classification on high-dimensional data and reported that the higher-order of the adaptive Lasso method outperformed on large dispersion, but the higher-order of the adaptive elastic net method was preferred with small dispersion. Recently, Sudjai et al[20] compared the performances of the power of the adaptive Lasso and adaptive elastic net methods under high-dimensional sparse data with multicollinearity. Their results showed that the square root of the adaptive elastic net performed best. However, there are no studies that have compared how penalized methods perform in high-dimensional sparse data with multicollinearity, emphasizing the initial weight together with the power of adaptive weight for the adaptive Lasso. Choosing appropriate the initial weight and the power of adaptive weight maybe help to improved efficacy of variable selection procedure. In our case study, if adjusting the initial weight together with the power of adaptive weight could be help to improve the efficacy of variable selection procedure, the adaptive Lasso can exactly extract significant information and subsequently produce the classification model that has the potential for differentiating benign adipocytic tumors from adipocytic malignancies.

Therefore, we examined whether adjusting the initial weight and the power of adaptive weight for the adaptive Lasso could be help to improve the efficacy of variable selection procedure. The primary objective of this study was to compare the performances of the initial weight together with the power of adaptive weight for the adaptive Lasso under high-dimensional sparse data with multicollinearity in a simulation study. Additionally, the classification performance of these approaches was compared on an adipocytic tumors data application.

Materials and Methods

Logistic regression model

The logistic regression is one of most supervised ML algorithms employed for classification tasks when the aim is to predict the probability that an instance belongs to a given class or not. In the cases of the binary outcome variable, a dependent variable (Y_i) would simply be coded as 1 (positive class) or 0 (negative class),which has a Bernoulli distribution ($Y_i \sim Bernoulli(\pi_i)$) with the parameter $\pi_i = \text{epx}(\underset{\sim}{x}_i\underset{\sim}{\beta}) / \left[1+\exp(\underset{\sim}{x}_i\underset{\sim}{\beta})\right]$. $\underset{\sim}{\beta} = \left(\beta_0, \beta_1, \beta_2, ..., \beta_p\right)^T$ presents a vector composed of logistic regression coefficients and $\underset{\sim}{x}_i = \left[1, x_{i1}, x_{i2}, ..., x_{ip}\right]$ denotes a vector of independent variables for the i^{th} observation where $i = 1,2,3,...,n$. n is the sample size and P is the number of independent variables. Therefore, the binary logistic regression model is as follows:

$$Y_i = \pi_i + \varepsilon_i,\ i = 1,2,3,...,n \tag{1}$$

where ε_i denotes the random error, which is assumed to follow a distribution with zero mean and variance of $\pi_i\left(1-\pi_i\right)$.

The transformation of π_i is a central of logistic regression (known as the logit function), which can be written as follows:

$$\ln\left(\frac{\pi_i}{1-\pi_i}\right) = \underset{\sim}{x}_i\underset{\sim}{\beta} = \beta_0 + \sum_{j=1}^{p} x_{ij}\beta_j \tag{2}$$

where $\underset{\sim}{\beta} = (\beta_0, \beta_1, \beta_2, ..., \beta_p)^T$; $i = 1,2,3,...,n$; and

$j = 1, 2, 3, ..., p$.

The log-likelihood function for a set of observations $(y_i, \underset{\sim}{x}_i)$ can be written as:

$$\ell(\underset{\sim}{\beta}) = \sum_{i=1}^{n} \left[\begin{array}{l} y_i \left(\beta_0 + \sum_{j=1}^{p} x_{ij}\beta_j \right) \\ - \ln\left(1 + \exp\left(\beta_0 + \sum_{j=1}^{p} x_{ij}\beta_j \right) \right) \end{array} \right]. \quad (3)$$

We can use the MLE (standard approach) to determine the estimated parameters of equation (3), which is as follows:

$$\hat{\underset{\sim}{\beta}}_{MLE} = \arg\max_{\underset{\sim}{\beta}} \left(\begin{array}{l} y_i \left(\beta_0 + \sum_{j=1}^{p} x_{ij}\beta_j \right) \\ - \ln\left(1 + \exp\left(\beta_0 + \sum_{j=1}^{p} x_{ij}\beta_j \right) \right) \end{array} \right) \quad (4)$$

where $\hat{\underset{\sim}{\beta}}_{MLE}$ represents a $(p+1)\times 1$ vector of the maximum likelihood estimators. Nevertheless, this method has some limitation with regards to multicollinearity and high-dimensional data. Especially, multicollinearity is a problem that affects the logistic regression models in which two or more independent variables (also called predictor variables) are high correlated. When this arises, the maximum likelihood estimators of the logistic regression coefficients trends to be very imprecise, that is, it has inflated variance, even if the sample size is sufficiently large. Ignoring this problem can cause in unreliable estimated coefficients (i.e., the estimates very sensitive to minor changes in the model or also called unstable), making it difficult to determine the actual effect of each predictor and could lead to misinterpretations of the significance of predictors. Therefore, dealing multicolleariy is necessary to ensure that the estimated coefficients reflect the actual relationship between the predictor and outcome variables. Thus, the penalized logistic regression is used as an alternative to the MLE. From equation (3), we can be written in a form of penalized approach as follows:

$$\ell^*(\underset{\sim}{\beta}) = -\ell(\underset{\sim}{\beta}) + P_\lambda(\underset{\sim}{\beta}) \quad (5)$$

where $P_\lambda(\underset{\sim}{\beta})$ denotes the penalty function and λ represents the tuning parameter.

Penalized logistic regression model

In cancer research, an overfitted model is highly accurate when predicting patients from its own data set, but has poor predictive performance when applied to patients in other data sets. Thus, shrinkage term in the penalized logistic regression model is designed to prevent extreme values of logistic regression coefficients, reducing the risk of overfitting in model development. Additionally, it is advantageous when the sample size is small relative to the number of predictors and multicollinearity arises. The penalized logistic regression differs from the MLE by incorporating a shrinkage term into the model. This model is given below.

$$\hat{\underset{\sim}{\beta}}_{PLR} = \arg\min_{\underset{\sim}{\beta}} \left(-\left\{ \begin{array}{l} y_i \left(\beta_0 + \sum_{j=1}^{p} x_{ij}\beta_j \right) \\ - \ln\left(1 + \exp\left(\beta_0 + \sum_{j=1}^{p} x_{ij}\beta_j \right) \right) \end{array} \right\} + P_\lambda(\underset{\sim}{\beta}) \right) \; ; \lambda \geq 0 \quad (6)$$

where $P_\lambda(\underset{\sim}{\beta})$ is term of a penalty function. λ represents a tuning parameter. If λ equals zero, $\hat{\underset{\sim}{\beta}}_{PLR} = \hat{\underset{\sim}{\beta}}_{MLE}$. For selecting the optimal value of λ, cross-validation is commonly utilized to evaluate the value of this parameter.

Shrinkage penalty is popular approach to alleviate the overfitting phenomenon by shrinking the logistic regression coefficients toward zero and moves extreme predicted values toward the average risk, leading to more accurate predictions when the model is applied in new patients. Different types of shrinkage penalties depending on the form of penalty function. Presently, the Lasso and adaptive Lasso are popular methods for cancer classification,[3, 21-23] which are described below.

Lasso

Lasso was originally designed to alleviate the disadvantage of ridge regression (i.e., the inability to reduce the number of independent variables in the final model). This approach proposed by Tibshirani in 1996.[15] The Lasso coefficient estimates, just like in ridge regression, are shrunk towards zero. The Lasso penalty (ℓ_1 -norm penalty) is defined by:

$$P_\lambda^{lasso}(\underset{\sim}{\beta}) = \lambda \sum_{j=1}^{p} \left|\beta_j\right|. \tag{7}$$

Thus, the coefficient estimates using the Lasso penalty is defined as follows:

$$\hat{\underset{\sim}{\beta}}_{lasso} = \arg\min_{\underset{\sim}{\beta}} \left(-\left\{ \begin{array}{l} y_i\left(\beta_0 + \sum_{j=1}^{p} x_{ij}\beta_j\right) \\ -\ln\left(1+\exp\left(\beta_0 + \sum_{j=1}^{p} x_{ij}\beta_j\right)\right) \end{array} \right\} + P_\lambda^{lasso}(\underset{\sim}{\beta}) \right) \; ; \lambda > 0 \; . \tag{8}$$

We use cross-validation strategies to evaluate the optimal value of tuning parameter λ. This parameter controls the shrinkage of $\hat{\underset{\sim}{\beta}}$. [24, 25] When the tuning parameter is sufficiently large, the Lasso penalty will shrink some coefficient estimates to exactly zero. This means that Lasso has the capability to perform variable selection. Consequently, the model derived from Lasso is simpler to interpret than that from ridge regression.[26] Despite being able to handle high-dimensional data, Lasso still has its limitations. First, if $p \gg n$, Lasso selects at most n independent variables before it saturates. Moreover, if a group of variables has high correlation, Lasso will choose only one variable and doesn't prioritize which one.[16] Finally, Lasso lacks oracle properties.[17, 27]

Adaptive Lasso

The lack of oracle properties is a significant reason why Lasso may exhibit instability.[27] Thus, in 2006, Zou[17] proposed adaptive Lasso to overcome this drawback. The principle of adaptive Lasso is a different weight for parameter $\underset{\sim}{\beta}$ in ℓ_1 -norm penalty. The adaptive Lasso penalty is given below.

$$P_\lambda^{Alasso}(\underset{\sim}{\beta}) = \lambda \sum_{j=1}^{p} w_j \left|\beta_j\right|. \tag{9}$$

Therefore, the estimation of $\underset{\sim}{\beta}$ using the adaptive Lasso penalty can be determined as follows:

$$\hat{\underset{\sim}{\beta}}_{Alasso} = \arg\min_{\underset{\sim}{\beta}} \left(-\left\{ \begin{array}{l} y_i\left(\beta_0 + \sum_{j=1}^{p} x_{ij}\beta_j\right) \\ -\ln\left(1+\exp\left(\beta_0 + \sum_{j=1}^{p} x_{ij}\beta_j\right)\right) \end{array} \right\} + P_\lambda^{Alasso}(\underset{\sim}{\beta}) \right) \tag{10}$$

where $w = (w_1, w_2, ..., w_p)^T$ represents a vector composed of weight vector and $w_j = \left|\hat{\beta}_j\right|^{-\gamma}$; $\gamma > 0$. γ denotes the power of adaptive weight. We can notice that w_j depends on the root n-consistent initial values of $\hat{\beta}_j$. The initial weight can be determined by using the MLE/ridge regression/Lasso.[12, 17] To reduce selection bias, this weighted approach assigns smaller weights to large coefficients and larger weights to small coefficients. Therefore, the adaptive Lasso can truly enjoy oracle properties.[17]

Apart from the above penalty functions, ridge regression, elastic net, and adaptive elastic net are also being increasingly employed in cancer research in data sets in which the target population is not rare diseases.[28-30] Regarding selecting specific penalty functions, we compare the advantages, limitations, and suitability of applying each shrinkage penalty in the penalized logistic regression model as shown in Table 1.

Monte Carlo simulation

The number of predictors, sample size, and degree of correlation for the predictors affect classification model accuracy. In this simulation study, we focused on two conditions:

(1) High-dimensional sparse data ($p > n$).[8] Assuming a sparse coefficients, we defined the number of significant predictors as q, and $q < p$ $\underset{\sim}{x}_i = \left(\underset{\sim}{x}_{iA}, \underset{\sim}{x}_{iB}\right)$ with $\underset{\sim}{x}_{iA} = \left(x_{i1}, x_{i2}, x_{i3}, \ldots, x_{iq}\right)^T \in \mathbb{R}^q$ as well as $\underset{\sim}{x}_{iB} = \left(x_{i(q+1)}, x_{i(q+2)}, x_{i(q+3)}, \ldots, x_{ip}\right)^T \in \mathbb{R}^{p-q}$.

Thus, $\underset{\sim}{X} = \left(\underset{\sim}{x}_A, \underset{\sim}{x}_B\right)^T \in \mathbb{R}^{n \times p}$ is the matrix of all independent variables when $\underset{\sim}{x}_A = \left(\underset{\sim}{x}_{iA}, \ldots, \underset{\sim}{x}_{nA}\right)^T \in \mathbb{R}^{n \times q}$ and

Table 1. Advantages, limitations, and suitability of applying each shrinkage penalty in the penalized logistic regression model.

Penalty function	Advantages	Limitations	Suitability of application
Ridge	- Able to overcome the problem of multicollinearity - Able to deal with low-/high-dimensional data	- Lacks variable selection property	- Multicollinearity arises - The data sets are low-/high-dimensional. - All independent variables relate to the outcome variable.
Lasso	- Able to handles the multicollinearity problem - Lasso is one of the standard sparse methods, but it is non-robust.	- If the number of independent variables (p) are less than the sample size (n) and the independent variables are high correlated, Lasso is dominated by ridge regression - If $p > n$, Lasso selects a most n variables before it saturates. - When independent variables in the data set have a high pairwise correlation, Lasso selects only one/a few variables from the correlated group, regardless of which ones are chosen. - Lacks oracle properties	- The independent variables are low/medium correlated. - The data are high-dimensional.
Adaptive Lasso	- Performance of the adaptive Lasso surpasses that of the Lasso and elastic net. - The estimators have oracle properties	-	- The independent variables are highly collinearity. - The data are high-dimensional.
Elastic net	-	- Lacks oracle properties	- Multicollinearity exists. - The data are high-dimensional.
Adaptive elastic net	- The adaptive elastic net outperforms the elastic net. - The estimators have oracle properties	-	- The independent variables are highly correlated. - The data are high-dimensional.

$\underset{\sim}{x}_B = \left(\underset{\sim}{x}_{iB}, \ldots, \underset{\sim}{x}_{nB}\right)^T \in \mathbb{R}^{n\times(p-q)}$.

(2) All independent variables are collinear based on the Toeplitz correlation structure, which is as following.[31]

$$\sum_k = \begin{pmatrix} 1 & \rho & \rho^2 & \rho^3 & \cdots & \rho^{k-1} \\ \rho & 1 & \rho & \rho^2 & \cdots & \rho^{k-2} \\ \rho^2 & \rho & 1 & \rho & \cdots & \rho^{k-3} \\ \rho^3 & \rho^2 & \rho & 1 & \cdots & \rho^{k-4} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{k-1} & \rho^{k-2} & \rho^{k-3} & \rho^{k-4} & \cdots & 1 \end{pmatrix}_{k\times k} \quad (11)$$

where k is a positive integer that represents the number of the independent variables. Along with, let a value of ρ range from 0 to 1.

The Monte Carlo simulations were done using 50 independent variables (p). The sample size (n) was set to 30 and 40. We generated the dependent variables from the Bernoulli distribution with parameter π_i . The multivariate normal distribution with a mean of zero and covariance $\sum(\underset{\sim}{X} \sim N(0,\Sigma))$ was used to generate the independent variables. Degree of correlation (ρ) equaled 0.1, 0.3, 0.5, 0.75, 0.85, and 0.95. The logistic regression coefficients were assigned constant values.[19, 32] Along with, we take a sparse logistic regression parameter $\underset{\sim}{\beta}$: the proportion of nonzero components equals 50% or less of the total predictors p. Thus, the number of significant predictors (q) was set to 15.[8, 32] After that, the acquisition data was split into two subsets: 80% for the learning data set and 20% for the testing data set. Performances of the penalized methods was assessed using the predicted mean square errors (PMSE), which is the most popular statistic for regression issues. The PMSE is a crucial performance metric for quantifies prediction accuracy, which provides a measure of how well the model's predictions align with the actual data points. The lower value of the PMSE indicates better predictive accuracy.[33, 34] The evaluation of the PMSE is as follows:

$$\text{PMSE} = \sum_{i=1}^{n} \frac{(y_i - \hat{y}_i)^2}{n} \quad (12)$$

where y_i is i^{th} actual value of the dependent variable and $\hat{y}_i$ were i^{th} predicted value of the dependent variables. The optimal value of the tuning parameter (λ) was determined using 10-fold cross-validation strategy.[12, 16, 25] To product a stationary result, our experiment was repeated 1000 times. Thus, the MPMSE was determined from the mean of 1000 estimates of PMSE_j.

$$\text{MPMSE} = \frac{1}{1000} \sum_{j=1}^{1000} PMSE_j \ . \quad (13)$$

Penalized methods that produced the lowest MPMSE were regarded as the most favorable choice. Flowchart of the simulation procedure is shown in Figure 1.

Moreover, Figure 2 presents the workflow diagram of the machine-learning procedure for the real-data application. In this section, accuracy, recall, precision, F1-score, and the area under the receiver operator characteristic curve (AUC) were used as performance metrics for classification models.[33, 34] Each of them has values in the interval [0, 1]. The classification accuracy of each method was determined as following:

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN} \quad (14)$$

when TP, FP, TN, and FN denote true positive, false positive, true negative, and false negative, respectively (Table 2). Out of all instances, it counts how many were correctly classified.

For recall, we can be determined as follows:

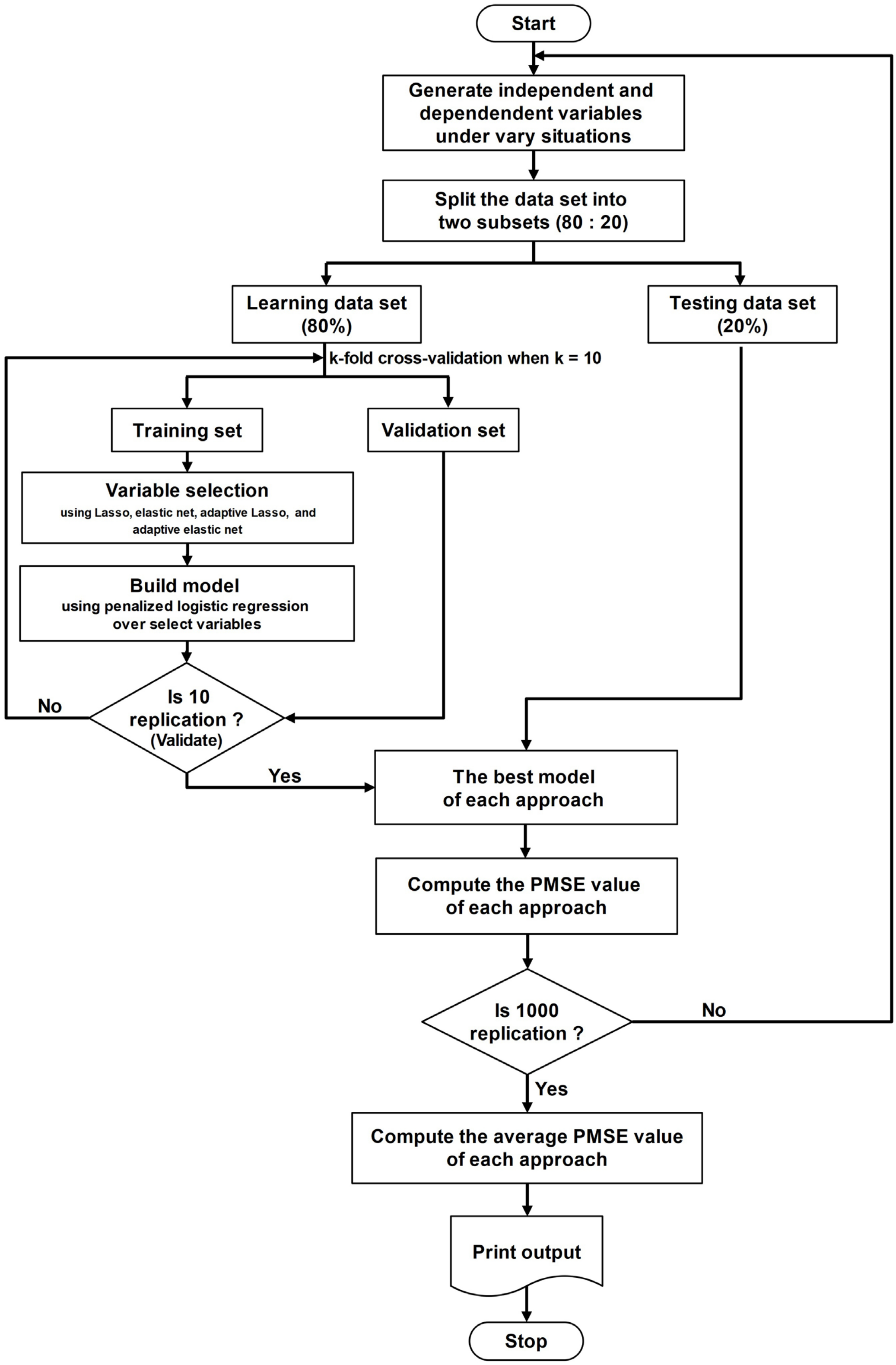

Figure 1. Flowchart of simulation procedure

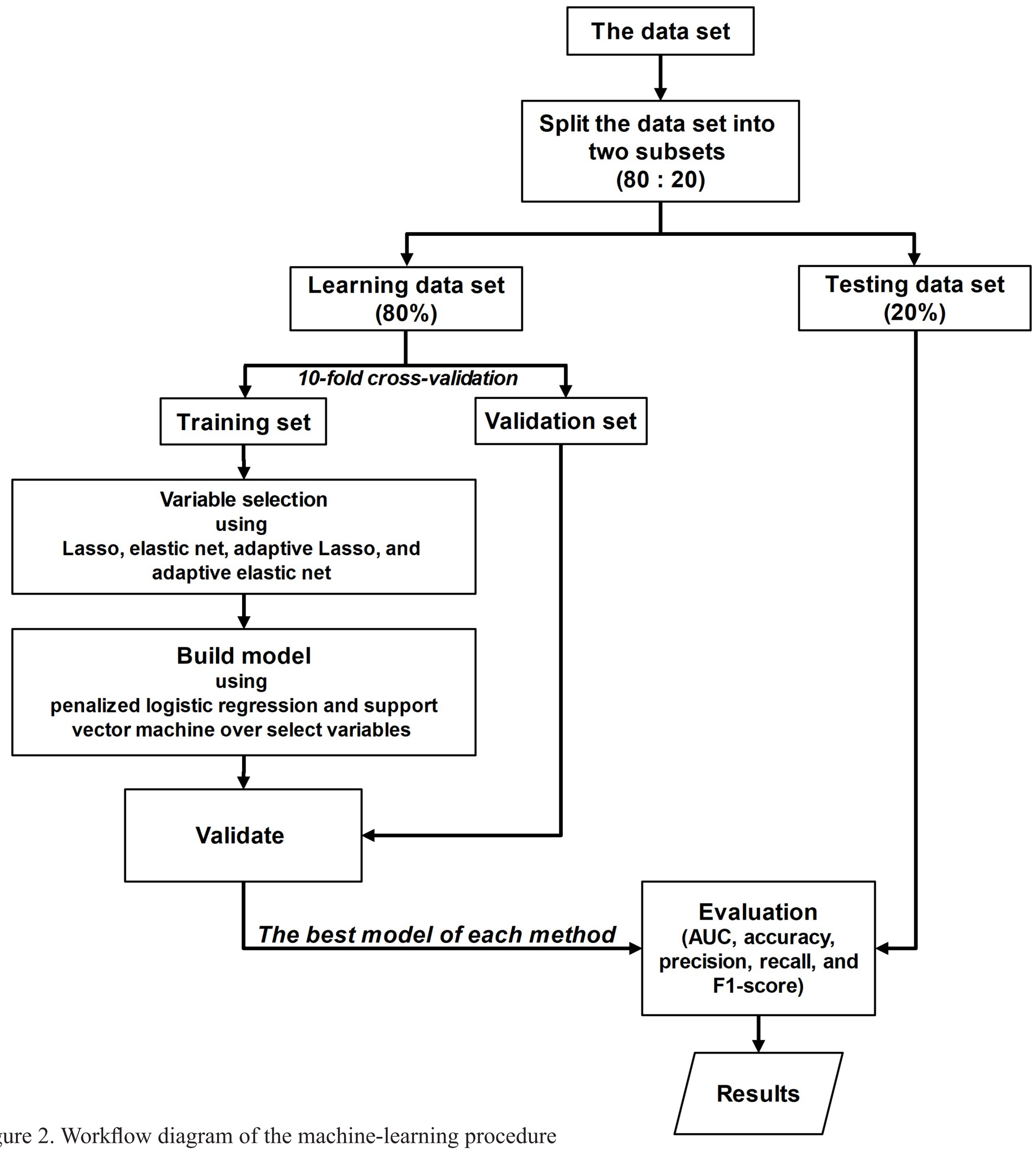

Figure 2. Workflow diagram of the machine-learning procedure

Table 2. Confusion matrix: Summary of prediction results on a classification task to appraise the performance of a machine learning model.

	Actual/Gold standard	
Predicted	Disease (number)	Non-disease (number)
Positive (number)	True positive (TP)	False positive (FP)
Negative (number)	False negative (FN)	True negative (TN)

$$\text{Recall} = \frac{TP}{TP+FN}. \quad (15)$$

Recall (also called sensitivity) is the ability of a predictive model to yield a positive result for a subject that has that disease. When recall is low, it denotes that a large number of disease cases are not detected.

Regarding precision and F1-score, precision is the proportion of positive predictions that were actually correct (also known as positive predictive value) and the F1-score is the harmonic mean of precision and recall that provided a balanced measure of performance). They have the formulas as follows:

$$\text{Precision} = \frac{TP}{TP+FP} \quad (16)$$

$$\text{and F1-score} = \frac{2\times\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}. \quad (17)$$

For AUC, with a value of 1 signifies that the ability of a predictive model to always correctly classify a randomly presented case (also called perfect classification). All of them were the most commonly utilized performance metrics of binary classification.[33]

Results

Simulation study

Table 3 shows the MPMSE values of each penalized method (i.e., Lasso, elastic net, adaptive Lasso, and adaptive elastic net). With high-dimensional sparse data (p = *50*, n = 30 and 40, and for different ρ), we found that the performance of the adaptive Lasso depended on the power of the adaptive weight (γ) and the initial weight used. In the case of $\rho = 0.1, 0.3,$ and 0.5, the MPMSE values of this method using $\underset{\sim}{w}_j = \left|\hat{\beta}_j^{ridge}\right|^{-1}$ were less than those

Table 3. Mean of the predicted mean square errors (MPMSE) values for the penalized methods when p = 50.

n	ρ	LASSO	Elastic net	Adaptive LASSO						Adaptive elastic net		
				$\underset{\sim}{w}_j = \left\|\hat{\beta}_j^{ridge}\right\|^{-\gamma}$			$\underset{\sim}{w}_j = \left\|\hat{\beta}_j^{lasso}\right\|^{-\gamma}$			$\underset{\sim}{w}_j = \left\|\hat{\beta}_j^{elasticnet}\right\|^{-\gamma}$		
				$\gamma=0.5$	$\gamma=1$	$\gamma=2$	$\gamma=0.5$	$\gamma=1$	$\gamma=2$	$\gamma=0.5$	$\gamma=1$	$\gamma=2$
30	0.10	0.177	0.180	0.166	0.162*	0.163	0.167	0.169	0.174	0.168	0.171	0.175
	0.30	0.177	0.181	0.168	0.165*	0.166	0.168	0.171	0.174	0.171	0.174	0.177
	0.50	0.183	0.188	0.173	0.171*	0.172	0.174	0.177	0.182	0.175	0.178	0.182
	0.75	0.189	0.192	0.185	0.184	0.185	0.183*	0.184	0.187	0.186	0.188	0.191
	0.85	0.192	0.198	0.191	0.191	0.191	0.188*	0.190	0.193	0.191	0.193	0.196
	0.95	0.196	0.198	0.195	0.195	0.198	0.192*	0.193	0.196	0.195	0.196	0.198
40	0.10	0.167	0.171	0.159	0.156*	0.157	0.157	0.161	0.165	0.158	0.161	0.166
	0.30	0.174	0.174	0.164	0.162*	0.164	0.165	0.167	0.172	0.164	0.167	0.172
	0.50	0.175	0.176	0.167	0.166*	0.168	0.168	0.170	0.174	0.167	0.170	0.174
	0.75	0.185	0.188	0.182	0.182	0.183	0.178*	0.180	0.184	0.180	0.183	0.186
	0.85	0.187	0.191	0.189	0.189	0.192	0.183*	0.186	0.189	0.186	0.188	0.193
	0.95	0.194	0.195	0.197	0.199	0.203	0.190*	0.192	0.195	0.193	0.194	0.197

Lasso, Least Absolute Shrinkage and Selection Operator; *The penalized methods providing the lowest MPMSE for the other methods. For $\rho = 0.75$, 0.85, and 0.95, the inflation of the MPMSE values for $\underset{\sim}{w}_j = \left|\hat{\beta}_j^{lasso}\right|^{-0.5}$ was the smallest compared with the other methods. Additionally, we compared the MPMSE values between four different approaches when $\underset{\sim}{w}_j$ for the adaptive Lasso is equal to $\left|\hat{\beta}_j^{lasso}\right|^{-0.05}$ and $\underset{\sim}{w}_j$ for the adaptive elastic net is equal to $\left|\hat{\beta}_j^{elasticnet}\right|^{-0.05}$ (Figure 3). The results showed that there was no statistically

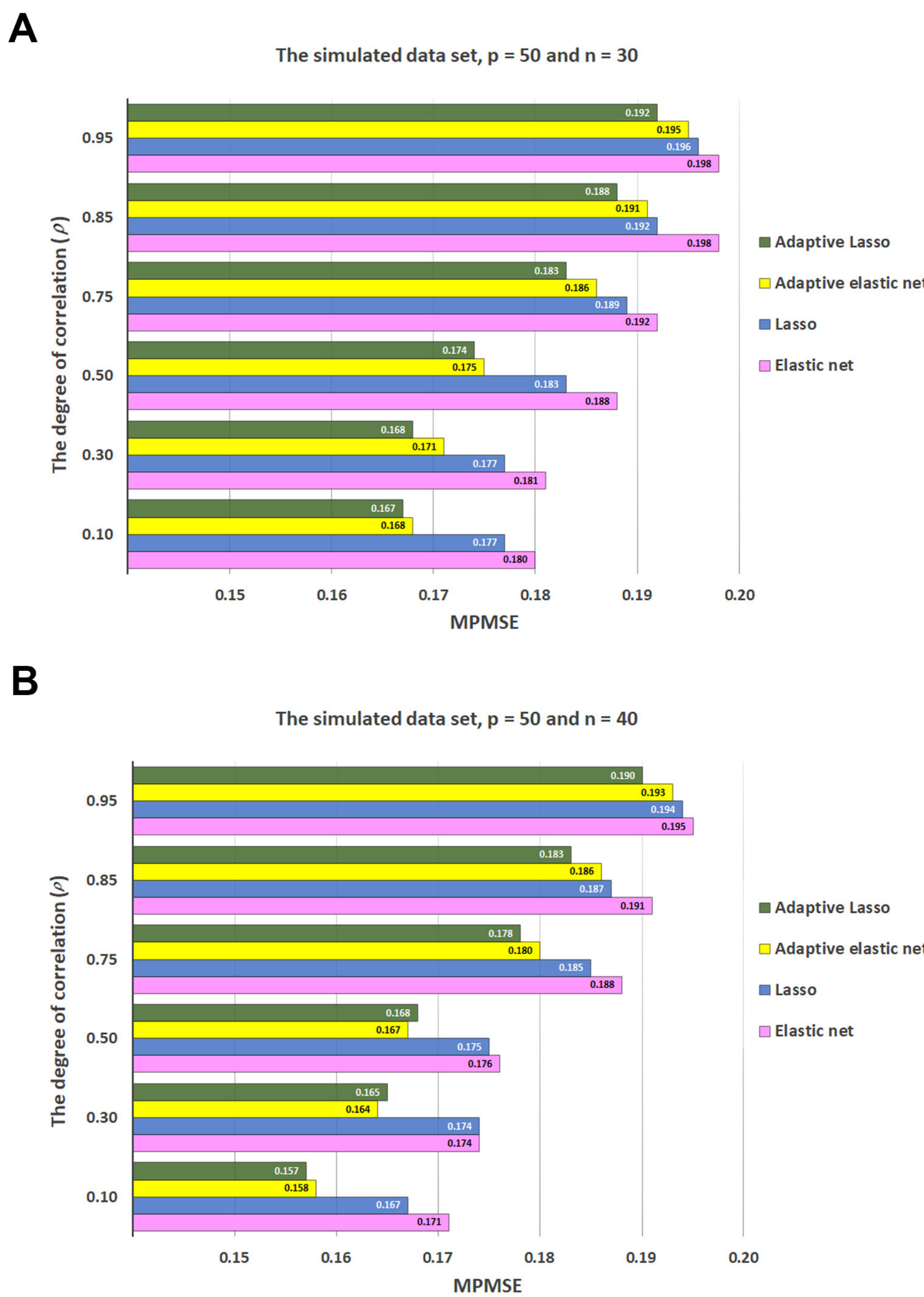

Figure 3. The mean of the predicted mean square errors (MPMSE) values of each penalized approach, which signified the performance of the approaches: A) The simulated data set: $p = 50$ and $n = 30$; and B) The simulated data set, $p = 50$ and $n = 40$. A lower value of MPMSE represents a better performance.

significant difference in MPMSE between methods (all $P > 0.05$ according to the Mann-Whitney U test).

Real-data applications

Presently, developments in ML algorithms have brought revolutionary changes to diagnostics in cancer such as adipocytic tumors. Most of the adipocytic tumors are either lipoma (benign tumor) and WDLS (malignant tumor), which frequently found in clinical practice. Distinguishing between IM lipomas that are symptomatic, deep-seated, and larger than 5 cm and WDLSs can be difficult due to their similarity. Thus, in our case study, we used MRI-based texture features to serve as non-invasive predictive biomarkers together with a machine-learning strategy for differentiating IM lipomas from WDLSs. Adipocytic tumors data set was obtained from 40 patients, which comprised 20 IM lipomas and 20 WDLSs. All patients underwent preoperative MRI scans including T1-weighted (T1W) sequence

Table 4. List of fifty MRI-based texture features

Gray-level co-occurrence matrix (GLCM)	Gray-level run length matrix (GLRLM)	Gray-level size zone matrix (GLSZM)	Neighbouring gray tone difference matrix (NGTDM)	Gray level dependence matrix (GLDM)
glcm_Autocorrelation (f1)	glrlm_ShortRun Emphasis (f17)	glszm_GrayLevel Non-Uniformity (f30)	ngtdm_Contrast (f43)	gldm_GrayLevelNon Uniformity (f44)
glcm_ClusterProminence (f2)	glrlm_RunPercentage (f18)	glszm_Zone Percentage (f31)		gldm_SmallDependence Emphasis (f45)
glcm_ClusterTendency (f3)	glrlm_RunEntropy (f19)	glszm_Zone Entropy (f32)		gldm_LargeDependence Emphasis (f46)
glcm_Contrast (f4)	glrlm_GrayLevelVariance (f20)	glszm_GrayLevelVariance (f33)		gldm_DependenceNonUniformity (f47)
glcm_Correlation (f5)	glrlm_HighGrayLevelRunEmphasis (f21)	glszm_HighGrayLevel-ZoneEmphasis (f34)		gldm_GrayLevel Variance (f48)
glcm_DifferenceAverage (f6)	glrlm_LongRunEmphasis (f22)	glszm_LargeAreaEmphasis (f35)		gldm_DependenceNonUniformity Normalized (f49)
glcm_DifferenceEntropy (f7)	glrlm_LongRunHighGray-LevelEmphasis (f23)	glszm_LargeAreaHigh-GrayLevelEmphasis (f36)		gldm_LowGrayLevel Emphasis (f50)
glcm_DifferenceVariance (f8)	glrlm_LongRunLowGray-LevelEmphasis (f24)	glszm_LargeAreaLow-GrayLevelEmphasis (f37)		
glcm_InverseVariance (f9)	glrlm_LowGrayLevelRunEmphasis (f25)	glszm_LowGrayLevel-ZoneEmphasis (f38)		
glcm_JointAverage (f10)	glrlm_RunLengthNonUniformity (f26)	glszm_SizeZoneNonUniformity (f39)		
glcm_JointEnergy (f11)	glrlm_RunLengthNonUniformity-Normalized (f27)	glszm_SmallAreaEmphasis (f40)		
glcm_JointEntropy (f12)	glrlm_ShortRunHighGray-LevelEmphasis (f28)	glszm_SmallAreaHigh-GrayLevelEmphasis (f41)		
glcm_MaximumProbability (f13)	glrlm_ShortRunLowGray-LevelEmphasis (f29)	glszm_SmallAreaLow-GrayLevelEmphasis (f42)		
glcm_SumAverage (f14)				
glcm_SumEntropy (f15)				
glcm_SumSquares (f16)				

Table 5. Example of the formula and definition of each MRI-based texture feature.

Feature class name	Feature name	Formula	Definition
Gray-level run-length matrix (GLRLM): *Where* N_g *is the number of discreet intensity values in the image* N_r *is the number of discreet run lengths in the image.* $P(i,j)$ *is the run-length matrix for an arbitrary direction* θ, *when* $i=1,2,3,...,N_g$ *and* $j=1,2,3,...,N_r$. N_p *is the number of voxels in the image*	glrlm_Short Run Emphasis (SRE)	$\mathrm{SRE}=\dfrac{\sum_{i=1}^{N_g}\sum_{j=1}^{N_r}\dfrac{P(i,j\|\theta)}{j^2}}{N_r(\theta)}$	Measures the distribution of short run lengths, with a greater value indicating shorter run lengths and more fine textural textures.
	glrlm_Run Percentage (RP)	$\mathrm{RP}=\dfrac{N_r(\theta)}{N_p}$	Measures the coarseness of the texture by taking the ratio of number of runs and number of voxels in the ROI, with higher values indicating more fine texture).
	glrlm_Run Entropy (RE)	$\mathrm{RE}=-\sum_{i=1}^{N_g}\sum_{j=1}^{N_r}P(i,j\|\theta)\log_2\left[P(i,j\|\theta)+\varepsilon\right]$	Measures the uncertainty/randomness in the distribution of run lengths and gray levels, with a higher value indicating more heterogeneity in the texture patterns.
Gray level size zone matrix (GLSZM): *Where* N_g *is the number of discreet intensity values in the image.* N_s *is the number of discreet zone sizes in the image.* N_p *is the number of voxels in the image.* N_z *is the number of zones in the ROI, which is equal to* $\sum_{i=1}^{N_g}\sum_{j=1}^{N_s}P(i,j)$ *and* $1\le N_z\le N_p$. $P(i,j)$ *is the size zone matrix, when* $i=1,2,3,...,N_g$ *and* $j=1,2,3,...,N_s$. $\varepsilon\approx 2.2\times10^{-16}$	glszm_Gray Level Non-Uniformity (GLN)	$\mathrm{GLN}=\dfrac{\sum_{i=1}^{N_g}\left(\sum_{j=1}^{N_s}P(i,j)\right)^2}{N_z}$	Measures the variability of gray-level intensity values in the image, with a lower value signifying more homogeneity in intensity values.
	glszm_Zone Percentage (ZP)	$\mathrm{ZP}=\dfrac{N_z}{N_p}$	Measures the coarseness of the texture by taking the ratio of number of zones and number of voxels in the ROI, with higher values indicating more fine texture.
	glszm_Zone Entropy (ZE)	$\mathrm{ZE}=-\sum_{i=1}^{N_g}\sum_{j=1}^{N_s}P(i,j)\log_2\left[P(i,j)+\varepsilon\right]$	Measures the uncertainty/randomness in the distribution of zone sizes and gray levels, with a higher value indicating more heterogeneneity in the

and total excision surgery at our institution. Subsequently, their diagnosis was confirmed using final pathological findings between 2010 and 2023. The predictors of interest were 50 texture features as continuous variables, which were extracted from MRI T1W images. These MRI-based texture features obtained from 5 classes: (1) gray-level co-occurrence matrix (GLCM); (2) gray-level run-length matrix (GLRLM); (3) gray-level size-zone matrix (GLSZM); (4) neighboring gray-tone-difference matrix (NGTDM); and (5) gray-level dependence matrix (GLDM). Regarding formula and definition of the features were explained according to PyRadiomics' documentation (https://pyradiomics.readthedocs.io/en/latest/features.html, accessed on 22 October 2024) (Tables 4 and 5). The outcome of interest was either an IM lipoma or a WDLS. We can see that the

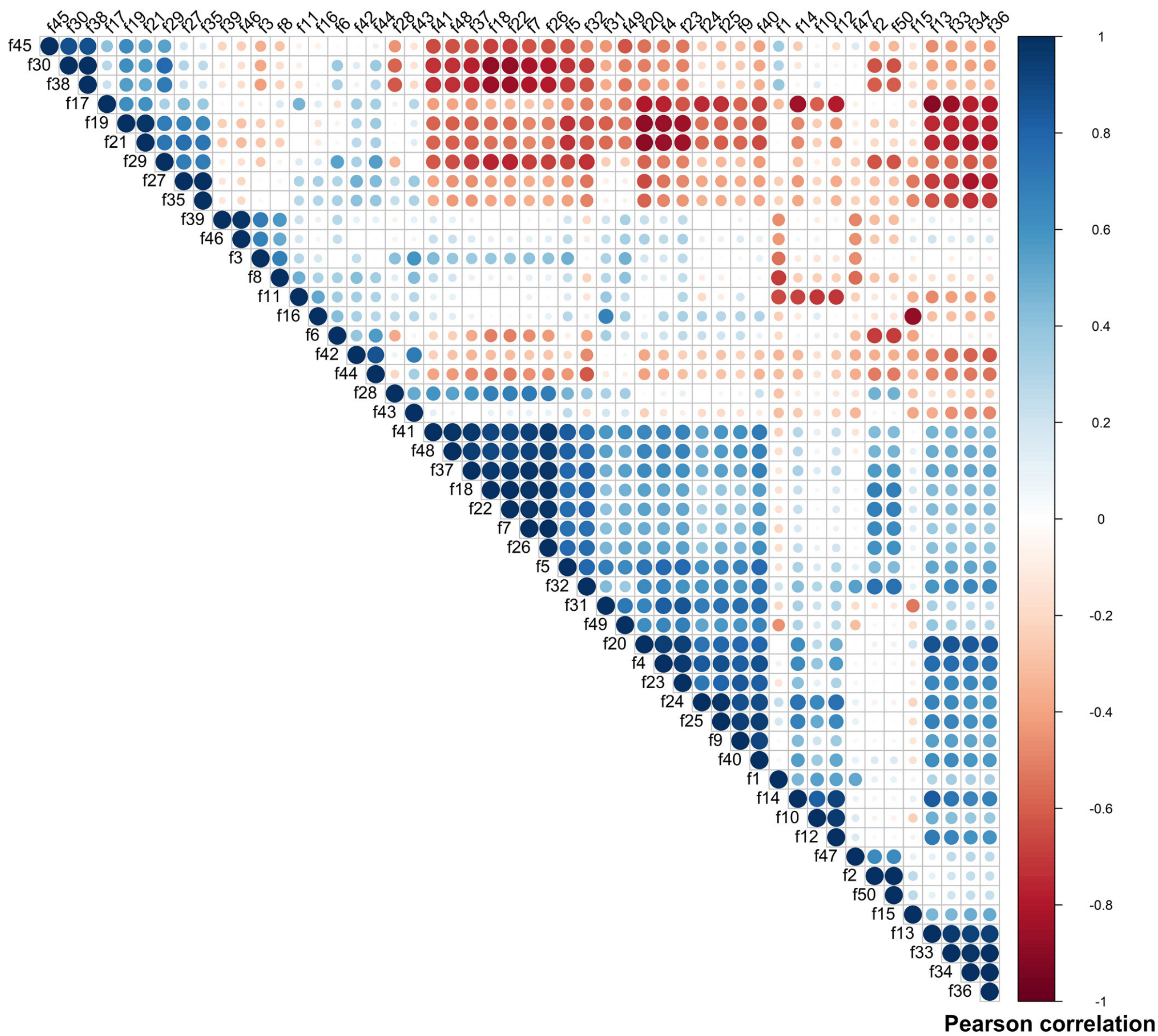

Figure 4. Correlation matrix of fifty texture features for adipocytic tumors data set, representing the different shades. The dark shade signifies a strong correlation among the features. Conversely, the light color indicates a weak correlation.

Table 6. Performances of machine-learning algorithms for distinguishing between intramuscular lipomas and well-differentiated liposarcomas for 50 texture features in 40 patients.

Classifier method	Variable selection approach										
	Lasso	Elastic net	Adaptive Lasso						Adaptive elastic net		
			$w_j = \lvert\hat{\beta}_j^{ridge}\rvert^{-\gamma}$			$w_j = \lvert\hat{\beta}_j^{lasso}\rvert^{-\gamma}$			$w_j = \lvert\hat{\beta}_j^{elasticnet}\rvert^{-\gamma}$		
			$\gamma = 0.5$	$\gamma = 1$	$\gamma = 2$	$\gamma = 0.5$	$\gamma = 1$	$\gamma = 2$	$\gamma = 0.5$	$\gamma = 1$	$\gamma = 2$
Penalized LR											
AUC	0.888	0.865	0.920	0.900	0.897	0.935	0.917	0.900	0.923	0.910	0.899
Accuracy	0.860	0.840	0.905	0.887	0.870	0.928	0.890	0.883	0.915	0.895	0.880
Precision	0.857	0.838	0.900	0.880	0.866	0.919	0.887	0.879	0.910	0.890	0.876
Recall	0.859	0.838	0.902	0.880	0.867	0.921	0.888	0.879	0.911	0.890	0.878
F1-score	0.857	0.836	0.900	0.880	0.868	0.925	0.888	0.880	0.910	0.891	0.877
SVM											
AUC	0.890	0.888	0.930	0.925	0.910	0.946	0.920	0.914	0.921	0.910	0.884
Accuracy	0.885	0.870	0.925	0.910	0.890	0.935	0.912	0.900	0.911	0.890	0.876
Precision	0.880	0.866	0.920	0.912	0.889	0.932	0.909	0.881	0.908	0.883	0.871
Recall	0.880	0.867	0.922	0.912	0.886	0.934	0.910	0.885	0.905	0.883	0.872
F1-score	0.875	0.862	0.920	0.910	0.885	0.930	0.905	0.880	0.905	0.880	0.867

Lasso, Least Absolute Shrinkage and Selection Operator; Penalized LR, penalized logistic regression; SVM, support vector machine; AUC, the area under the receiver operator characteristic curve.

number of predictors was more than the sample size. It indicates that the high-dimensional problem was presented. Additionally, Figure 4 shows the correlation matrix for the adipocytic tumors data set. The light shade denotes that the predictors have a low correlation, while the dark shade presents a high correlation among predictors. Furthermore, almost predictors had the variance inflation factor (VIF) values exceeding 10, indicating a concerning level of collinearity. Therefore, it is clear that the presence of multicollinearity is evident in this data. These above problems can cause the variance of logistic regression coefficient estimators to inflate and model overfitting. To solve all these problems, we employed the penalized methods (i.e., Lasso, elastic net, adaptive Lasso, and adaptive elastic net) on this case study. The classification performances of these penalized approaches are detailed in Table 6. In the cases of classifier with penalized logistic regression, we can see that the highest accuracy value (0.928) was derived from the adaptive Lasso with $w_j = \lvert\hat{\beta}_j^{lasso}\rvert^{-0.5}$, whereas the lowest accuracy value (0.840) was derived from the elastic net method. Their respective precisions, recalls, F1-scores, and AUCs were 0.919, 0.921, 0.925, and 0.935, respectively for this adaptive Lasso and 0.838, 0.838, 0.836, and 0.865, respectively for the elastic net. For classifier with SVM, the highest accuracy value (0.935) was also derived from the adaptive Lasso with $w_j = \lvert\hat{\beta}_j^{lasso}\rvert^{-0.5}$, whereas the lowest accuracy value (0.870) was also derived from the elastic net method. Their respective precisions, recalls, F1-scores, and AUCs were

0.932, 0.934, 0.930, and 0.946, respectively for this adaptive Lasso and 0.866, 0.867, 0.862, and 0.888, respectively for the elastic net.

Discussion

This study revealed that the significant factors effecting the performance of penalized approaches under high-dimensional sparse data with multicollinearity were the power of adaptive weight and initial weight. In Table 3, we can see that the adaptive Lasso produced the smallest MPMSE. The key factors that impacted the MPMSE included the power of adaptive weight, initial weight, the degree of correlation, and sample size. When ρ was increased while holding p and n fixed, the MPMSE values of all penalized approaches increased. Along with, the values of this MPMSE reduced with an increase in n. The worst outcome was obtained with a high degree of correlation (ρ = 0.95). For power of the adaptive weight on the ℓ_1-norm penalty, choosing $\underset{\sim}{w}_j = \left|\hat{\beta}_j^{lasso}\right|^{-0.5}$ is preferred for high-dimensional sparse data with multicollinearity. Although there was no statistically significant difference in the MPMSE values between methods, the estimated coefficient values in adaptive Lasso with $\underset{\sim}{w}_j = \left|\hat{\beta}_j^{lasso}\right|^{-0.5}$ exhibited the lowest bias from the actual values (Figure 3). In Table 6, the results of the real-data applications evident that the adaptive Lasso has superior performance to the other methods. We can see that the results of the simulation study support this finding.

Two previous studies compared the performances of the power of the adaptive Lasso and adaptive elastic net methods. First, Araveeporn[19] compared the higher-order of the adaptive Lasso and adaptive elastic net methods for classification on high-dimensional data. Their independent variables were generated from the normal distribution with small and large dispersions. The results showed that the higher-order of the adaptive Lasso method outperformed on large dispersion, but the higher-order of the adaptive elastic net method was preferred with small dispersion. For another study, Sudjai et al[20] reported that the square root of the adaptive elastic net performed best under high-dimensional sparse data with multicollinearity. The results of our study are comparable with those of these studies. However, both two previous studies, there are no consideration present in choosing appropriate the initial weight for the adaptive Lasso approach. Our study demonstrated that adjusting the initial weight together with the power of adaptive weight for the adaptive Lasso can enhance the efficiency of the variable selection process, leading to the construction of a classification model that can the potentially distinguish IM lipomas from WDLSs.

Regarding a confounding factor in the real-data applications, a potential confounders were controlled by setting the following inclusion criteria as follows: (1) the final pathological diagnosis must be confirmed as IM lipoma or WDLSs; (2) received surgical excision of primary tumor; (3) received preoperative T1W MRI protocol. The exclusion criteria were set as follows: (1) the patients underwent anticancer treatments before undergoing the MRI scan; (2) having a history of other malignant tumors; (3) poor quality MRI image.

Our study has some limitations. First, we did not explore all possible data structures. Consequently, we were not able to identify all estimation difficulties that may arise in modeling real life data sets. Thus, we can only generalize our findings to a real-data set that has the problems of high-dimensional sparse data with multicollinearity. Another limitation, we did not explore modeling with qualitative (categorical) covariates. Therefore, the guidelines of this article should only be considered for quantitative (continuous) predictors.

Conclusion

Adjusting the initial weight together with the power of adaptive weight can help improve the efficacy of variable selection procedure, i.e., the adaptive Lasso can exactly extract significant information, and subsequently produce the classification model that has the potential for differentiating benign adipocytic tumors from adipocytic malignancies. We suggest the use of the adaptive Lasso method for classification in binary outcome on high-dimensional sparse data with multicollinearity as follows: we should be considered using the first power of the method (i.e., $\gamma = 1$) and ridge estimator as the initial weight when there is a low or moderate correlation among the independent variables. In contrast, if the independent variables are high collinear, the initial weight should be evaluated using the Lasso estimator and $\gamma = 0.5$. In practice, implementing above recommendation is usefulness as follows: (1) it may help to lessen diagnostic uncertainty in medical image classification with MRI; (2) it helps in alerting physicians to promptly refer a patient to a sarcoma center when needed; (3) reduce unnecessary biopsy; and (4) assists surgeons decide whether to hold off or expedite surgery. A further study should be developed the ML algorithms and modifying the mathematical approaches that will increase the efficacy of classification in diagnostics of cancers under all possible data structures for simulation study and real-world scenarios.

Author Contributions

Conceptualization, N.S., M.D., C.C.; methodology, N.S., M.D., C.C.; software, N.S., M.D.; validation, N.S., M.D., C.C.; formal analysis, N.S., M.D., C.C.; investigation, N.S., C.C.; resources, C.C.; data curation, N.S., C.C.; writing—original draft preparation, N.S.; writing—review and editing, N.S., M.D., C.C.; project administration, N.S.; funding acquisition, C.C. All authors finally read and approved the manuscript.

Funding

The study was supported by the Siriraj Foundation Fund for advanced sarcoma research (grant number D004146).

Institutional Review Board statement

The adipocytic The adipocytic tumors data set in this study was retrieved after approval from the Ethics Committee of the Faculty of Medicine Siriraj Hospital, Mahidol University (approval number MU-MOU CoA No. 874/2020).

Acknowledgments

We thank Supani Duangkaew for her research coordination and Pakorn Yodprom for providing physics consultation and retrieving magnetic resonance images.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Makalic E, Schmidt DF. Review of modern logistic regression methods with application to small and medium sample size problems. In: Li J, ed. AI 2010: Advances in Artificial Intelligence. Berlin, Heidelberg: Springer 2011:213-22.

2. Sudjai N, Duangsaphon M. Liu-type logistic regression coefficient estimation with multicollinearity using the bootstrapping method. Science, Engineering and Health Studies. 2020;14(3):203-14.

3. Sudjai N, Siriwanarangsun P, Lektrakul N, Saiviroonporn P, Maungsomboon S, Phimolsarnti R, et al. Tumor-to-bone distance and radiomic features on MRI distinguish intramuscular lipomas from well-differentiated liposarcomas. J Orthop Surg Res. 2023;18(1):255.

4. Sudjai N, Siriwanarangsun P, Lektrakul N, Saiviroonporn P, Maungsomboon S, Phimolsarnti R, et al. Robustness of radiomic features: two-dimensional versus three-dimensional MRI-based feature reproducibility in lipomatous soft-tissue tumors. Diagnostics. 2023;13(2):258.

5. Hosmer DW, Lemeshow SJ. Applied logistic regression. 3 ed. New Jersey: Wiley; 2013.

6. Kleinbaum DG, Klein M. Logistic regression: a self-learning text. 3rd ed. New York: Springer; 2010.

7. Senaviratna NAMR, Cooray TMJA. Multicollinearity in binary logistic regression model. Theory and Practice of Mathematics and Computer Science. 2021;8: p. 11-9.

8. Cherkassky V, Mulier F. Learning from data: concepts, theory, and methods. 2nd ed. New Jersey: John Wiley and Sons; 2006.

9. Brimacombe M. High-dimensional data and linear models: a review. Open Access Med Stat. 2014;4:17-27.

10. Belsley DA, Kuh E, Welsch RE. Regression diagnostics: Identifying influential data and sources of collinearity. New York: John Wiley & Sons; 1980.

11. Kastrin A, Peterlin B. Rasch-based high-dimensionality data reduction and class prediction with applications to microarray gene expression data. Expert Syst Appl. 2010;37:5178-85.

12. Pavlou M, Ambler G, Seaman S, De Iorio M, Omar RZ. Review and evaluation of penalised regression methods for risk prediction in low-dimensional data with few events. Stat Med. 2016;35(7):1159-77.

13. Hosseinnataj A, Bahrampour A, Baneshi M, Zolala F, Nikbakht R, Torabi M, et al. Penalized Lasso methods in health data: application to trauma and influenza data of Kerman. Journal of Kerman University of Medical Sciences. 2019;26(6):440-9.

14. Hoerl AE, Kennard RW. Ridge regression: biased estimation for nonorthogonal problems. Technometrics. 1970;12(1):55-67.

15. Tibshirani R. Regression shrinkage and selection via the Lasso. J R Stat Soc Series B Stat Methodol. 1996;58(1):267-88.

16. Zou H, Hastie T. Regularization and variable selection via the elastic Net. J R Stat Soc Series B Stat Methodol. 2005;67(2):301-20.

17. Zou H. The adaptive Lasso and Its oracle properties. J Am Stat Assoc. 2006;101(476):1418-29.

18. Zou H, Zhang HH. On the adaptive elastic-net with a diverging number of parameters. Ann Stat. 2009;37(4):1733-51.

19. Araveeporn A. The higher-order of adaptive Lasso and elastic net methods for classification on high dimensional data. Mathematics. 2021;9(10):1091.

20. Sudjai N, Duangsaphon M, Chandhanayingyong C. Adaptive elastic net on high-dimensional sparse data with multicollinearity: Application to lipomatous tumor classification. International Journal of Statistics in Medical Research. 2024;13:30-40.

21. Foreman SC, Llorián-Salvador O, David DE, Rösner VKN, Rischewski JF, Feuerriegel GC, et al. Development and evaluation of MR-Based radiogenomic models to differentiate atypical lipomatous tumors from lipomas. Cancers (Basel). 2023;15(7):2150.

22. Piffoux M, Jacquemin J, Pétéra M, Durand S, Abila A, Centeno D, et al. Metabolomic prediction of breast cancer treatment induced neurological and metabolic toxicities. Clin Cancer Res. 2024;30(20):4654-66.

23. Wang L, Zhao Z, Luo Y, Yu H, Wu S, Ren X, et al. Classifying 2-year recurrence in patients with dlbcl using clinical variables with imbalanced data and machine learning methods. Comput Methods Programs Biomed. 2020;196:105567.

24. Efron B, Hastie T, Johnstone I, Tibshirani R. Least angle regression. Ann Stat. 2004;32(2):407-51.

25. Hastie T, Tibshirani T, Friedman JB. The Elements of statistical learning: data mining inference and prediction. 2nd ed. Berlin/Heidelberg: Springer; 2009.

26. James G, Witten D, Hastie T, Tibshirani R. An Introduction to statistical learning with applications in R. New York: Springer; 2013.

27. Fan J, Li R. Variable selection via nonconcave penalized likelihood and its oracle properties. J Am Stat Assoc. 2001;96(456):1348-60.

28. Kamalapathy PN, Ramkumar DB, Karhade AV, Kelly S, Raskin K, Schwab J, et al. Development of machine learning model algorithm for prediction of 5-year soft tissue myxoid liposarcoma survival. J Surg Oncol 2021;123(7):1610-7.

29. Kamalapathy PN, Gonzalez MR, de Groot TM, Ramkumar D, Raskin KA, Ashkani-

Esfahani S, et al. Prediction of 5-year survival in soft tissue leiomyosarcoma using a machine learning model algorithm. J Surg Oncol 2024;129(3):531-6.

30. Toğaçar M, Ergen B, Cömert Z. Application of breast cancer diagnosis based on a combination of convolutional neural networks, ridge regression and linear discriminant analysis using invasive breast cancer images processed with autoencoders. Med Hypotheses. 2020;135:109503.

31. Hardin J, Garcia SR, Golan D. A method for generating realistic correlation matrices. Ann Appl Stat. 2013;7(3):1733-62, 30.

32. Bottmer L, Croux C, Wilms I. Sparse regression for large data sets with outliers. Eur J Oper Res. 2022;297(2):782-94.

33. Rainio O, Teuho J, Klén R. Evaluation metrics and statistical tests for machine learning. Sci Rep. 2024;14(1):6086.

34. Tohka J, van Gils M. Evaluation of machine learning algorithms for health and wellness applications: A tutorial. Comput Biol Med. 2021;132:104324.