Electronic Journal of Applied Statistical Analysis
EJASA, Electron. J. App. Stat. Anal.
http://siba-ese.unisalento.it/index.php/ejasa/index
e-ISSN: 2070-5948
DOI: 10.1285/i20705948v16n3p507

Classification of Breast Cancer Histopathological Images using Adaptive Penalized Logistic Regression with Wilcoxon Rank Sum Test
By Alkahya, Alreahan and Algamal

15 December 2023

Electronic Journal of Applied Statistical Analysis
Vol. 16, Issue 3, 2023, 507-518
DOI: 10.1285/i20705948v16n3p507

Classification of Breast Cancer Histopathological Images using Adaptive Penalized Logistic Regression with Wilcoxon Rank Sum Test

Mohammed Abdulrazaq Alkahya*[a], Hussein Obeid Alreahan[b], and Zakariya Yahya Algamal[c]

[a] *Department of Computer science, College of Education for Pure Science, University of Mosul, Mosul, Iraq*
[b] *Department of Mathematics, College of Education for Pure Science, University of Mosul, Mosul, Iraq*
[c] *Department of Statistics and Informatics, College of Computer science and Mathematics, University of Mosul, Mosul, Iraq*

15 December 2023

Classification of histopathological images is crucial in diagnosis and treatment. Selecting the most helpful image features from thousands of candidates remains a critical challenge to classify these images. In this paper, a compatible methodology is proposed to identify features by combining logistic regression with weighted L1-norm. Our study tested our proposed method and compared it with state-of-the-art methods using publicly available breast cancer histopathological image datasets. Our results indicate that our method outperforms three other methods regarding overall classification accuracy and the number of selected features.

keywords: feature selection; lasso; penalized logistic regression; breast cancer; histopathological image; Wilcoxon rank sum test.

1 Introduction

According to the World Health Organization (WHO), breast cancer is one of the leading causes of cancer-related deaths in women worldwide (Singh et al., 2016). However, early

*Corresponding author: Tel.: +964 7701656690, mohammedkahya@uomosul.edu.iq.

ISSN: 2070-5948
http://siba-ese.unisalento.it/index.php/ejasa/index

diagnosis can significantly increase survival rates (Sudarshan et al., 2016). Various non-invasive imaging methods such as mammography (X-rays), magnetic resonance imaging (MRI), and ultrasonography, are available (Acharya et al., 2016; Spanhol et al., 2015). Recently, histopathological image analysis has become a significant research issue in medical imaging in recent times (Vu et al., 2015). Breast cancer diagnosis via histopathology images is the gold standard, helping to narrow borderline diagnoses from other standard imaging (Zhang et al., 2013). The use of machine learning has improved the diagnostic accuracy of breast cancer in women when embedded in a computer-aided system (Belsare et al., 2015). In breast cancer classification, identifying normal and abnormal cell patterns is crucial for diagnosis and drug discovery (Korkmaz et al., 2014; Zhang et al., 2015). Improving patient healthcare is crucial, and accurate cancer prediction is invaluable for treatment (Chen et al., 2016; Mao et al., 2013).

Logistic regression is a commonly used classification method in various fields, particularly in the classification of medical, commercial, weather, and gene expression data. (Kahya, 2019; Algamal, 2017a). As the number of features or genes increases, the computational complexity and training time of logistic regression also increase (Algamal, 2017a; Algamal and Lee, 2018). Although logistic regression is advantageous in gene expression data classification, it cannot handle gene selection automatically (Qasim and Algamal, 2018; Algamal, 2017b; Kahya et al., 2017). Penalized methods are very effective embedded gene selection methods connected with many popular classification methods. In recent years, logistic regression among all the classification methods, those based on sparseness, received much attention. It combines the logistic regression with a penalty to simultaneously perform gene selection and classification. With different penalties, several logistic regression models can be applied, among which are L1-norm, which is called the least absolute shrinkage and selection operator (lasso) (Tibshirani, 1996), smoothly clipped absolute deviation (SCAD) (Fan and Li, 2001), elastic net(Zou and Hastie, 2005), and adaptive L1-norm(Zou, 2006). Undoubtedly, L1-norm is one of the most commonly used gene sparsity methods. However, this method results in inconsistent gene selection due to the same amount of sparseness applied to all genes (Algamal and Lee, 2015; Zou, 2006). In this study, we propose an efficient method for selecting informative features and classifying breast cancer histopathological images. Our approach combines penalized logistic regression with the Wilcoxon rank sum test. We use the Wilcoxon rank sum test to weigh each feature and assign a weight to each significant feature based on its test value. This weight reflects the importance of each feature. We also use a penalized logistic regression with adaptive L1-norm to refine the feature selection process further. Overall, our method offers an effective way to select and classify features for breast cancer histopathological images.

Our proposed method is compared to competitor methods using the BreaKHis database, which consists of microscopic biopsy images of benign and malignant breast tumors (Spanhol et al., 2015).

The remainder of this paper is structured as follows: Section 2 provides an overview of penalized logistic regression and feature extraction. Section 3 outlines the proposed method and related procedures. In Section 4, details of the experimental study are presented. Section 5 presents the experimental results. Finally, in Section 6, general

conclusions are drawn.

2 Overview

2.1 Penalized Logistic Regression

Logistic regression is a statistical approach that models a binary classification problem. The regression function has a non-linear relationship with the linear combination of genes. In cancer classification, the response variable of the logistic regression takes on two values: 1 for the tumor class or 0 for the normal class.

Let $y_i \in \{0, 1\}$ be a vector of size $n \times 1$ of tissues, and let $\mathbf{x}_i$ is a $1 \times p$ vector of features. The logistic transformation of the vector of probability estimates $\pi_i = p(y_i = 1 \mid x_i)$ is modeled by a linear function, logit transformation:

$$\ln[\frac{\pi_i}{1-\pi_i}] = \beta_0 + \sum_{j=1}^{p} \mathbf{x}_{ij}^T \beta_{\mathbf{j}}, \quad i = 1, 2, ..., n, \tag{1}$$

where β_0 is the intercept and β_j is a $p \times 1$ vector of unknown feature coefficients. The log-likelihood function of Eq.(1) is defined as:

$$\ell(\beta_{\mathbf{0}}, \beta) = \sum_{i=1}^{n} \{y_i \ln(\pi_i) + (1 - y_i) \ln(1 - \pi_i)\}. \tag{2}$$

Logistic regression offers the advantage of simultaneously estimating the probabilities π and $\pi - 1$ for each class and classifying subjects. The probability of classifying the i^{th} sample in class 1 is estimated by $\widehat{\pi}_i = \exp(\beta_0 + \sum_{j=1}^{p} \mathbf{x}_{ij}^T \beta_{\mathbf{j}})/(1 + \exp(\beta_0 + \sum_{j=1}^{p} \mathbf{x}_{ij}^T \beta_{\mathbf{j}}))$. The predicted class is then obtained by $I\{\widehat{\pi}_i > 0.5\}$, where $I(\cdot)$ is an indicator function. PLR adds a nonnegative penalty term to Eq.(1), such that the size of the gene coefficients in high-dimension cancer classification can be controlled. Several penalty terms have been discussed in the literature (Tibshirani, 1996; Liu et al., 2007; Hoerl and Kennard, 1970).The L_1- penalty, proposed by Tibshirani (Tibshirani, 1996) is one of the popular penalty terms. The L_1- penalty performs genes selection and stimation simultaneously by constraining the log-likelihood function of gene coefficients. The penalized method for the logistic regression is obtained by adding the penalty term to the negative log-likelihood function:

$$PLR = -\sum_{i=1}^{n} \{y_i \ln(\pi_i) + (1 - y_i) \ln(1 - \pi_i)\} + \lambda P(\beta). \tag{3}$$

The estimation of the vector β is obtained by minimizing Eq.(3):

$$\hat{\beta}_{PLR} = \arg\min_{\beta} \left[-\sum_{i=1}^{n} \{y_i \ln(\pi_i) + (1 - y_i) \ln(1 - \pi_i)\} + \lambda P(\beta) \right], \tag{4}$$

where $\lambda P(\beta)$) is the penalty term that penalizes the estimates. The penalty term depends on the positive tuning parameter λ , which controls the tradeoff between fitting

the data to the model and the effect of the penalty. In other words, it controls the amount of shrinkage. For $\lambda = 0$ we obtain the MLE solution, while for large values of λ the influence of the penalty term on the coefficient estimates increases. Choosing the tuning parameter is an important part of the model fitting. If we interesting in classification, the tuning parameter should find the right balance between the bias and the variance to minimize the misclassification error. Without loss of generality, it is assumed that the genes are standardized, $\sum_{i=1}^{n} \mathbf{x}_{ij} = 0$ and $(-1)\sum_{i=1}^{n} \mathbf{x}_{ij}^2 = 1 \quad \forall j \in \{i = 1, 2, ..., p\}$. As a result, the intercept β_0 is not penalized. The estimation of the vector β using LASSO (L_1-penalty) is defined as:

$$\hat{\beta}_{PLR} = \arg\min_{\beta} \left[-\sum_{i=1}^{n} \{y_i \ln(\pi_i) + (1 - y_i) \ln(1 - \pi_i)\} + \lambda \sum_{j=1}^{P} |\beta| \right], \tag{5}$$

where λ is a tuning parameter. It reduces to the MLE estimator when $\lambda = 0$. On the other hand, if $\lambda \to \infty$, , the penalty forces all the genes to be zeros. In practice, the value of λ is often chosen by a cross validation procedure. Eq.(5) can be efficiently solved by using the coordinate descent algorithm (Park and Hastie, 2007; Friedman et al., 2010).

The LASSO has an advantage in that it is computationally feasible in high dimensional classification data. On the other hand, the LASSO has three main drawbacks. First of all, if $p > n$, the LASSO selects at most n genes because of the nature of the convex optimization problem. In addition, the LASSO cannot handle the effect of grouping. When the pairwise correlations among a group of genes are very high, then the LASSO tends to select only one gene from the whole group and does not take into account which one is selected (Zou and Hastie, 2005; Feng et al., 2012). Lastly, the LASSO lacks the oracle properties, as stated by Fan and Li (Fan and Li, 2001).

2.2 Features Extraction

Features of breast cancer histopathological images were extracted using Haar discrete wavelet transform based on 7th-level decomposition (Balaji, 2007; Misiti, 2000; Gonzalez, 2008). Breast cancer histopathology images were initially divided into four equal-sized sub-images, namely LL1 (approximation coefficients), LH1 (horizontal coefficients), HL1 (vertical coefficient), and HH1 (diagonal coefficient), at the first level of decomposition. Subsequently, the LL1 (approximation coefficient) sub-image was further decomposed into four equal-sized sub-images, namely LL2, LH2, HL2, and HH2, at the second level of decomposition. This process was continued until the seventh level of decomposition was reached, forming 28 sub-images from every channel (red, green, and blue). Therefore, 28 x 3 sub-images were obtained from the original image.

Next, nine traditional statistical features were calculated, including mean, standard deviation, skewness, kurtosis, entropy, energy, root mean square, mean absolute deviation, and median absolute deviation. This resulted in the extraction of 756 features for each breast cancer histopathological image.

3 Proposed Method

Feature selection plays a vital role in the problem of classifying breast cancer images. It helps to enhance the accuracy of classification by minimizing the negative impact of irrelevant image features on the classification process. Therefore, it is important to carefully select the most relevant features to improve the accuracy of image classification and correct diagnosis. It is crucial to remove irrelevant and noisy features from the original feature matrix to efficiently apply classification techniques. Too many features can negatively impact the classifier's performance, resulting in overfitting and longer computational times.

In this paper, we present an efficient feature selection method that addresses the consistency problem in feature selection. Our approach tackles the issue of inconsistency that occurs when using PLR with L1-norm. This problem can lead to the selection of irrelevant features, which reduces the effectiveness of the method. To address the issue, we employed a combination of PLR and L1-norm techniques along with a Wilcoxon rank sum test. This test is utilized to evaluate the significance of each feature. Subsequently, we used PLR with adaptive L1-norm to assign weights to significant features based on their corresponding Wilcoxon rank sum test values. The weight assigned to each feature reflects its relative importance.

3.1 Weight Calculation

In practical scenarios, the feature matrix may contain irrelevant or noisy features, resulting in low performance and lower classification accuracy. As a result, it has become essential to analyze the features in terms of their importance. The Wilcoxon rank sum test can be used to determine each feature's weight (Liao et al., 2006) as:

$$S(j) = \sum_{i \in N_1} \sum_{k \in N_2} I((x_i^{(j)} - x_k^{(j)} \leq 0), \quad j = 1, 2, ..., p, \tag{6}$$

where $I(\cdot)$ is the discrimination function and it is defined as:

$$I(\cdot) = \begin{cases} 1 & \text{if } I \text{ is true} \\ 0 & \text{if } I \text{ is not true} \end{cases} \tag{7}$$

$x_i^{(j)}$ is the value of the sample i in the feature j, moreover N_1 and N_2 are the index sets of different classes of samples. Eq.(6), $S(j)$ represents the measurement of the difference between the two classes. The feature j can be considered important when Eq.(6) is close to 0 or when it is close to the max value of $n_1\ n_2$, where $n_1 = |N_1|$ and $n_2 = |N_2|$. Liao, Li and Luo (Liao et al., 2006) the feature significance by the following feature ranking criterion:

$$q(j) = \max\{S(j), n_1\ n_2 - S(j)\}. \tag{8}$$

Depending on Eq.(8), an important feature, with $S(j)$ closed to 0 or to $n_1\ n_2$, will receive large value of $q(j)$, while an irrelevant feature will receive a small value of $q(j)$.

To enforce discriminative penalty on each feature according to importance degree in classification, Park and other (Park et al., 2016) proposed the following weight as below:

$$w_i = 1/[\frac{q(j)}{\sum_{j=1}^{p} q(j)} \times p], \quad j = 1, 2, ..., p. \tag{9}$$

According to Eq.(9),the important feature will receive small amount of weight, while the irrelevant feature will receive relatively large amount of weight. By this weighting procedure, the L_1-norm can reduce the inconsistent property in feature selection.

3.2 Breast Cancer Classification

The process involves assigning a weight to each feature and then using the adaptive L_1-norm PLR to select the most informative features with high accuracy in classification. The Algorithm 1 describes the detailed computation of the adaptive PLR (APLR). The APLR equation takes on a convex form, which guarantees the existence of a global maximum point and makes it easy to solve efficiently.

Algorithm 1: The computation of APLR

Step 1: Find $w_j, \quad j = 1, 2, ..., p.$

Step 2: Define $\tilde{\mathbf{x}}_i = w_j \mathbf{x}_i$

Step 3: Solve the APLR,

$$\hat{\beta}_{\text{APLR}} = \arg\min_{\beta} \left\{ -\sum_{i=1}^{n} \{y_i \ln(\pi_i) + (1 - y_i) \ln(1 - \pi_i)\} + \lambda \sum_{j=1}^{p} w_{\text{j}} \left|\beta_{\text{j}}\right| \right\}$$

4 Experimental Study

4.1 Dataset description

The dataset used was The BreaKHis database, it is a comprehensive collection of high-resolution microscopic images depicting various types of breast tumors, including both benign and malignant ones. The database is a valuable resource for researchers and medical professionals working toward the early detection, diagnosis, and treatment of breast cancer (Spanhol et al., 2015). BreaKHis is a valuable dataset that originates from the pathological anatomy and cytopathology laboratory of Parana, Brazil. The dataset consists of 7,909 high-quality microscopic images of breast tumor tissue that have been carefully selected to be clinically representative. These images have been obtained from a diverse group of 82 patients and captured at varying magnifications, including 40X, 100X, 200X, and 400X. The dataset includes 24 benign and 58 malignant samples, making it a comprehensive resource for breast cancer research. For more details, refer to Table 1, which provides a detailed dataset summary.

Table 1: Summary of the BreaKHis database.

Magnification	Benign	Malignant
40X	625	1370
100X	644	1437
200X	623	1390
400X	588	1232

4.2 Performance Evaluation

To assess the effectiveness of the proposed method, two performance metrics have been implemented, namely the patient classification rate (PCR) and overall classification accuracy (OCA). The PCR represents the proportion of correctly classified benign and malignant cases within each patient. It can be defined as follows:

$$PCR = \frac{n_{correct}}{n_{total}} \times 100, \tag{10}$$

where $n_{correct}$ is the number of correctly classified cancer images for the patient j and n_{total} is the number of cancer images of patient j.

The OCA can define as:

$$OCA = \frac{\sum_{j=1}^{p} PCR_j}{n_{patients}} \times 100, \tag{11}$$

where $n_{patients}$ is the number of patients. Generally, the closer value to 1, the better overall classification performance is.

4.3 Experimental Setting

To demonstrate the effectiveness of the proposed method, we conducted comprehensive comparative experiments with the PLR-lasso, PLR-SCAD, and the classical LR. To do this, we randomly partitioned the data matrix into training and test datasets, where 70% of the samples were used for the training dataset and the remaining 30% for the testing dataset. For a fair comparison and to minimize the impact of the data partition, all the classification methods used were evaluated for their classification performance metrics using 10-fold cross-validation, averaged over 50 partitioned times.

Based on the training dataset, the tuning parameter value, lambda, was fixed for each classification method used as: $0 \leq \lambda \leq 100$. As suggested by Fan and Li, the SCAD penalty constant α was set to 3.7 (Fan and Li, 2001). The R-package "penalized LR" contains the necessary implementations for these methods.

5 Experimental Results

5.1 Classification Performance

Table 2 summarizes, on average, the overall classification accuracy for the training and testing datasets of applying the APLR, PLR-SCAD, PLR-lasso, and the LR. In addition, it summarizes the number of the selected features. The number in parenthesis is the corresponding standard deviation. Beginning with the magnification 40X, regarding the overall classification accuracy and based on the training dataset, the proposed method, APLR, achieves 95.71%, defeating PLR-SCAD, PLR-lasso, and the LR by 5.27%, 8.01%, and 12.95%, respectively. In addition, PLR-SCAD secondly comes with 90.44% and better than PLR-lasso and LR. Depending on the testing dataset, the APLR is better than the others in terms of overall classification accuracy because it achieved 93.98%, which is 5.52%, 9.52%, and 13.95% better than PLR-SCAD, PLR-lasso, and the LR, respectively. In the magnification 100X, based on the training dataset, the APLR provides enhancement over the PLR-SCAD and the PLR-lasso by 5.02% and 7.43%, respectively. Once again, based on the testing dataset, the proposed method beats both PLR-SCAD and PLR-lasso in terms of overall classification accuracy. Looking at the magnification 200X, the overall classification performance of the proposed method is comparable with PLR-SCAD, PLR-lasso, and LR performing best among them. In terms of overall classification accuracy, the OCA obtained from the proposed method was 97.58% for the training dataset and 95.04% for the testing dataset. This indicates the superiority of the proposed method as compared to PLR-SCAD, PLR-lasso, and LR. At the end, regarding the magnification 400X, the APLR shows a considerable dominance against PLR-SCAD, PLR-lasso, and LR. It achieved the higher overall classification accuracy for both the training and testing datasets. PLR-SCAD and PLR-lasso attain useful performance as they provide results that are inferior to APLR but better than LR. The number of features selected by each method is an important factor. Methods selecting more features tend to overfit the data. Hence, methods with a small number of selected features are preferred. For a comparison of methods in terms of the number of selected features, the APLR outperformed PLR-SCAD and PLR-lasso. For instance, in magnification 200X, APLR selected 6 features compared to 13 and 20 features for the PLR-SCAD and PLR-lasso, respectively.

5.2 Statistical Significance Test

To further confirm the effectiveness of the proposed feature selection method in achieving high classification performance, a pairwise comparison was made using a two-tailed T-test between the proposed method and each competitor method. The comparison was based on the area under the curve of the training dataset. Figure 1 displays the boxplot of the AUC for the different methods used. It is evident from the graph that the proposed method's AUC is comparable to the results obtained from PLR-SCAD and PLR-Lasso. Table 3 shows the results of paired two-tailed t-tests at a significance level of $\alpha = 0.05$. The AUC of the proposed method is statistically significantly better than that of PLR-SCAD, PLR-lasso, and LR.

Table 2: Classification performance of the APLR, PLR-SCAD, PLR-lasso and LR.

Magnification	Method	Training dataset	Testing dataset	#features
40x	APLR	95.71 (1.109)	93.98 (2.983)	9
	PLR-SCAD	90.44 (1.457)	88.46 (2.567)	18
	PLR-lasso	87.71 (2.911)	84.82 (4.021)	25
	LR	82.76 (2.943)	80.03 (3.127)	all
100x	APLR	96.25 (1.089)	94.31 (1.754)	8
	PLR-SCAD	91.23 (1.843)	88.56 (2.585)	14
	PLR-lasso	88.82 (2.899)	85.68 (3.832)	22
	LR	84.98 (2.132)	81.01 (4.049)	all
200x	APLR	97.58 (0.928)	95.04 (1.225)	6
	PLR-SCAD	92.35 (1.881)	89.85 (2.107)	13
	PLR-lasso	89.58 (2.543)	87.20 (4.027)	20
	LR	85.23 (2.136)	81.27 (3.854)	all
400x	APLR	94.65 (1.832)	92.21 (2.539)	11
	PLR-SCAD	89.37 (2.019)	87.35 (3.071)	19
	PLR-lasso	86.36 (2.811)	83.67 (3.407)	27
	LR	81.59 (3.013)	78.97 (4.208)	all

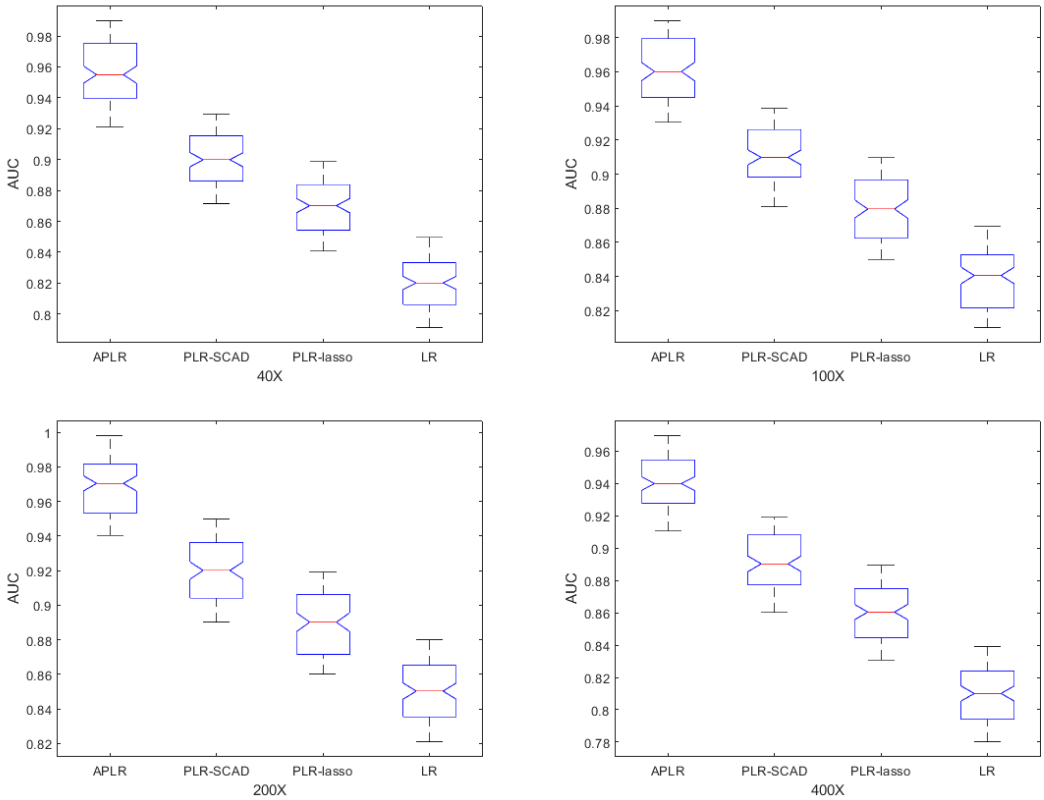

Figure 1: Boxplot of the AUC for all data magnifying : 40X, 100X, 200X, and 400X.

Table 3: P-values for the paired t-test of our proposed method results with competitor methods.

Magnification	APLR vs PLR-SCAD	APLR vs PLR-lasso	APLR vs PL
40x	0.0022(*)	0.0013(*)	0.0009(*)
100x	0.0041(*)	0.0035(*)	0.0015(*)
200x	0.0047(*)	0.0035(*)	0.0017(*)
400x	0.0009(*)	0.0005(*)	0.0002(*)

(*) means that the methods have significant differences

6 Conclusion

This paper proposes an adaptive penalized logistic regression model that combines the logistic regression model with the weighted L1-norm to classify breast cancer histopathology images. Our proposed method was tested experimentally and compared to other existing methods. The superior classification performance of our approach was demonstrated through the overall classification accuracy and the T-test for the AUC. The proposed method showed outstanding performance compared to PLR-SCAD, PLR-lasso, and LR. Specifically, the OCA achieved a classification accuracy of 97.58% for the training data and 95.04% for the testing data. These results demonstrate the effectiveness of the proposed method in achieving high accuracy in overall classification.Additionally, compared to other methods, the proposed method performs exceptionally well regarding the number of selected features. For instance, in magnification 200X, APLR selected six features compared to 13 and 20 for the PLR-SCAD and PLR-lasso, respectively.

Consequently, the results confirm that the proposed method is a promising feature selection technique for medical image classification. The availability of medical images in this field is limited, making access difficult and posing significant limitations for study and other future work.

References

Acharya, U. R., Raghavendra, U., Fujita, H., Hagiwara, Y., Koh, J. E., Hong, T. J., Sudarshan, V. K., Vijayananthan, A., Yeong, C. H., Gudigar, A., et al. (2016). Automated characterization of fatty liver disease and cirrhosis using curvelet transform and entropy features extracted from ultrasound images. *Computers in biology and medicine*, 79:250–258.

Algamal, Z. (2017a). Classification of gene expression autism data based on adaptive penalized logistic regression. *Electronic Journal of Applied Statistical Analysis*, 10(2):561–571.

Algamal, Z. (2017b). An efficient gene selection method for high-dimensional microarray data based on sparse logistic regression. *Electronic Journal of Applied Statistical Analysis*, 10(1):242–256.

Algamal, Z. Y. and 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.

Algamal, Z. Y. and Lee, M. H. (2018). A two-stage sparse logistic regression for optimal gene selection in high-dimensional microarray data classification. *Advances in Data Analysis and Classification*, pages 1–19.

Balaji, B. (2007). *Rotation and scale invariant texture classification using log polar wavelet energy signatures*. PhD thesis.

Belsare, A., Mushrif, M., Pangarkar, M., and Meshram, N. (2015). Classification of breast cancer histopathology images using texture feature analysis. In *TENCON 2015-2015 IEEE Region 10 Conference*, pages 1–5. IEEE.

Chen, Y., Wang, L., Li, L., Zhang, H., and Yuan, Z. (2016). Informative gene selection and the direct classification of tumors based on relative simplicity. *BMC bioinformatics*, 17(1):44.

Fan, J. and Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American statistical Association*, 96(456):1348–1360.

Feng, Z. Z., Yang, X., Subedi, S., and McNicholas, P. D. (2012). The lasso and sparse least squares regression methods for snp selection in predicting quantitative traits. *IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB)*, 9(2):629–636.

Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of statistical software*, 33(1):1.

Gonzalez, R. C. (2008). Rew digital image processing.

Hoerl, A. E. and Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1):55–67.

Kahya, M. A. (2019). Classification enhancement of breast cancer histopathological image using penalized logistic regression. *Indonesian Journal of Electrical Engineering and Computer Science*, 13(1):405–410.

Kahya, M. A., Al-Hayani, W., and Algamal, Z. Y. (2017). Classification of breast cancer histopathology images based on adaptive sparse support vector machine. *Journal of Applied Mathematics and Bioinformatics*, 7(1):49.

Korkmaz, S., Zararsiz, G., and Goksuluk, D. (2014). Drug/nondrug classification using support vector machines with various feature selection strategies. *Computer methods and programs in biomedicine*, 117(2):51–60.

Liao, C., Li, S., and Luo, Z. (2006). Gene selection using wilcoxon rank sum test and support vector machine for cancer classification. In *International Conference on Computational and Information Science*, pages 57–66. Springer.

Liu, Z., Jiang, F., Tian, G., Wang, S., Sato, F., Meltzer, S. J., and Tan, M. (2007). Sparse

logistic regression with lp penalty for biomarker identification. *Statistical Applications in Genetics and Molecular Biology*, 6(1).

Mao, Z., Cai, W., and Shao, X. (2013). Selecting significant genes by randomization test for cancer classification using gene expression data. *Journal of biomedical informatics*, 46(4):594–601.

Misiti, M. (2000). *Wavelet Toolbox for Use with MATLAB: User's Guide; Version 2; Computation, Visualization, Programming*. MathWorks Incorporated.

Park, H., Shiraishi, Y., Imoto, S., and Miyano, S. (2016). A novel adaptive penalized logistic regression for uncovering biomarker associated with anti-cancer drug sensitivity. *IEEE/ACM transactions on computational biology and bioinformatics*, 14(4):771–782.

Park, M. Y. and Hastie, T. (2007). Penalized logistic regression for detecting gene interactions. *Biostatistics*, 9(1):30–50.

Qasim, O. S. and Algamal, Z. Y. (2018). Feature selection using particle swarm optimization-based logistic regression model. *Chemometrics and Intelligent Laboratory Systems*, 182:41–46.

Singh, B. K., Verma, K., and Thoke, A. (2016). Fuzzy cluster based neural network classifier for classifying breast tumors in ultrasound images. *Expert Systems with Applications*, 66:114–123.

Spanhol, F. A., Oliveira, L. S., Petitjean, C., and Heutte, L. (2015). A dataset for breast cancer histopathological image classification. *IEEE Transactions on Biomedical Engineering*, 63(7):1455–1462.

Sudarshan, V. K., Mookiah, M. R. K., Acharya, U. R., Chandran, V., Molinari, F., Fujita, H., and Ng, K. H. (2016). Application of wavelet techniques for cancer diagnosis using ultrasound images: A review. *Computers in biology and medicine*, 69:97–111.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288.

Vu, T. H., Mousavi, H. S., Monga, V., Rao, G., and Rao, U. A. (2015). Histopathological image classification using discriminative feature-oriented dictionary learning. *IEEE transactions on medical imaging*, 35(3):738–751.

Zhang, L., Qian, L., Ding, C., Zhou, W., and Li, F. (2015). Similarity-balanced discriminant neighbor embedding and its application to cancer classification based on gene expression data. *Computers in biology and medicine*, 64:236–245.

Zhang, Y., Zhang, B., and Lu, W. (2013). Breast cancer histological image classification with multiple features and random subspace classifier ensemble. In *Knowledge-Based Systems in Biomedicine and Computational Life Science*, pages 27–42. Springer.

Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American statistical association*, 101(476):1418–1429.

Zou, H. and 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.