

Improving the Diagnosis of Breast Cancer Using Regularized Logistic Regression with Adaptive Elastic Net

Aiedh Mrisi Alharthi^{1,2}, Muhammad Hisyam Lee^{1,*}, Zakariya Yahya Algamal³

¹Department of Mathematical Sciences, Universiti Teknologi Malaysia, Johor, Malaysia

²Department of Mathematics, Taif University, Taif, Saudi Arabia

³Department of Statistics and Informatics, University of Mosul, Mosul, Iraq

Received August 12, 2021; Revised September 12, 2021; Accepted October 27, 2021

Cite This Paper in the following Citation Styles

(a): [1] Aiedh Mrisi Alharthi, Muhammad Hisyam Lee, Zakariya Yahya Algamal, "Improving the Diagnosis of Breast Cancer Using Regularized Logistic Regression with Adaptive Elastic Net," *Universal Journal of Public Health*, Vol. 9, No. 5, pp. 317 - 323, 2021. DOI: 10.13189/ujph.2021.090514.

(b): Aiedh Mrisi Alharthi, Muhammad Hisyam Lee, Zakariya Yahya Algamal (2021). *Improving the Diagnosis of Breast Cancer Using Regularized Logistic Regression with Adaptive Elastic Net. Universal Journal of Public Health*, 9(5), 317 - 323. DOI: 10.13189/ujph.2021.090514.

Copyright©2021 by authors, all rights reserved. Authors agree that this article remains permanently open access under the terms of the Creative Commons Attribution License 4.0 International License

Abstract Early diagnosis of breast cancer helps improve the patient's chance of survival. Therefore, cancer classification and feature selection are important research topics in medicine and biology. Recently, the adaptive elastic net was used effectively for feature-based cancer classification, allowing simultaneous feature selection and feature coefficient estimation. The adaptive elastic net basically employed elastic net estimates as the initial weight. Nevertheless, the elastic net estimator is inconsistent and biased in selecting features. Therefore, the regularized logistic regression with the adaptive elastic net (RLRAEN) was used to handle the inconsistency problem by employing the adjusted variances of features as weights within the L_1 -regularization of the elastic net model. The proposed method was applied to the Wisconsin Breast Cancer dataset of the UCI repository and compared to the other existing penalized methods that were also applied to the same dataset. Based on the experimental study, the RLRAEN was more efficient in terms of feature selection and classification accuracy than the other competing methods. Therefore, it can be concluded that RLRAEN is a better method in breast cancer classification.

Keywords Adaptive Elastic Net, Breast Cancer, Feature Selection, Regularized Logistic Regression

1. Introduction

Breast cancer is the world's second leading cause of death among women from cancer. Furthermore, it is one of the deadliest diseases among women. Unfortunately, it spreads faster in the world more than any other cancer disease. Unless it is detected in its early stages, breast cancer can threaten life [1]. Therefore, early diagnosis of breast cancer helps improve the patient's chance of survival. There are many approaches that can be used to diagnose breast cancer using many techniques of machine learning to decide classification, analysis, and prediction [2].

Over the recent decades, researchers have developed a variety of feature selection techniques. These techniques are divided into three groups. The first group is filter approaches. It includes the most common feature selection techniques, in which each feature is evaluated individually, irrespective of how well it performs in the group. The second group is wrapper approaches. It evaluates the feature group selection process using a variety of algorithms. Even though wrapper techniques, such as "forward feature selection" and "backward feature elimination" are more effective in feature selection than filter methods, wrapper methods are computationally very expensive. The embedded methods are the third group, which incorporates the benefits of both the filter and

wrapper groups. It contains penalization techniques that can model and select features simultaneously [3]–[5].

"Wisconsin diagnostic breast cancer" (WDBC) dataset was obtained from a digitized image of a breast mass using a fine needle [6]. In the machine learning discipline, there are numerous soft computing approaches that have been used to analyze and classify WDBC. The authors of Kadam et al. [2] indicated that several methods were proposed by many researchers in order to predict breast cancer early. Then, these authors proposed feature ensemble methods based on the so-called Stacked Sparse Autoencoders and Softmmx Regression Model (FE-SSAE-SM model) to classify breast cancer into benign and malignant. The findings revealed that their proposed method gave rise to a useful classification model of breast cancer. Thus, the method is as efficient as other machine learning methods.

One of the most popular penalty-based regularization methods is penalized logistic regression. It is also used to classify and select features. With logistic regression, embedded feature selection, including classification and regularized techniques, are very successful. The logistic regression has received much interest in recent years to conduct both feature selection and classification simultaneously. This incorporates logistic regression with a penalty. With various penalties, a variety of logistic regression models may be employed. Among these penalties is the "Least Absolute Shrinkage and Selection Operator" (LASSO), which is based on L_1 -regularization [7]. The "Smoothly Clipped Absolute Deviation" (SCAD) [8] is another penalty. Besides, the elastic net [9], the adaptive L_1 -regularization [10], and adaptive elastic net techniques [11], [12] are some of the other penalties.

The L_1 -regularization (LASSO) is capable of choosing variables. However, it has three flaws [13], [14]. The first flaw is linked to the number of features chosen by LASSO. The number of features in specific datasets may be much more than the number of observations. Regrettably, LASSO cannot choose more features than the number of observations. In other words, the number of features selected by LASSO is limited by the number of observations. The second flaw is linked to how the features operate. Basically, features operate as clusters or groupings. Each category contains features that are strongly correlated. This is anticipated to be taken into consideration by LASSO when choosing features. That is, it is anticipated to either pick the whole set of strongly associated features (assuming they are really associated with the illness) or leave it all alone (if they are unrelated). Regrettably, LASSO only chooses a feature of each highly correlated group of features relevant to the research. Zou and Hastie [9] developed the elastic net regularization technique to address the first and second limitations. The elastic net technique uses a penalty that is comprised of L_1 -regularization and L_2 -regularization. A bias in feature selection is considered the third flaw since it penalizes all

feature coefficients equally. Therefore, LASSO does not enjoy oracle properties [8]. In order to address this issue, Zou [10] devised a novel regularization method named the adaptive LASSO technique, in which different weights are employed to punish each coefficient within the L_1 -regularization penalty. Adaptive weights are used in the adaptive LASSO to punish various coefficients in the L_1 -regularization.

The L_1 -regularization is a popular technique in sparse approaches. One shortcoming of the L_1 -regularization model is that it applies the same amount of the penalty to all features, resulting in inconsistency in the feature selection process [8], [10]. A regularized logistic regression with adaptive elastic net (RLRAEN) is used in this research to enhance feature selection effectiveness. This is performed by using the adjusted variances of features as an initial weight within the L_1 -regularization with the elastic net to properly classify individuals in terms of catching cancer. This weight reflects the importance of each individual feature in certain respects. Experiments are carried out to compare the proposed feature selection technique in this study with other competitor methods.

The remainder of the paper is organized as follows. Section 2 gives a short overview of relevant research on regularized logistic regression techniques. Section 3 introduces the proposed method (RLRAEN). Section 4 presents and discusses the findings of the experimental research designed to assess the efficiency of RLRAEN in comparison to LASSO, elastic net, and adaptive elastic net. Eventually, Section 5 draws conclusions.

2. Regularized Logistic Regression

The logistic regression is used to model binary outcome variables. In the cancer classification issue, for example, the outcome variable has just two values: 1 for malignant tumours and 0 for non-cancerous tumours. The relationship between the regression equation and the linear combination of the predictor variables is nonlinear in the logistic regression.

To classify the outcome variable, \mathbf{y} , we assume that \mathbf{y} is the n -dimensional vector, where each of its components, y_i , equals 0 or 1. Classification of \mathbf{y} is based on a linear combination of the $p \times n$ matrix with real entries; designed as $\mathbf{X}^T \boldsymbol{\beta}$; where \mathbf{X}^T denotes the transpose of $\mathbf{X} = \{x_{ij}, 1 \leq i \leq n, 1 \leq j \leq p\}$ an $n \times p$ matrix of features, where x_{ij} takes its values in R , and $\boldsymbol{\beta}$ is the p -dimensional vector of unknown coefficients. In general, in the logistic regression, the outcome variable y has a Bernoulli distribution as (2), and the probability that y is equal to 1 given the value of x , indicated as

$\pi(x)$, is written as

$$p(y_i = 1 | x_{ij}) = \pi(x_i) = \frac{\exp(x_j^T \beta_j)}{1 + \exp(x_j^T \beta_j)}, j = 1, 2, \dots, p \quad (1)$$

$$f(y_i) = \pi_i^{y_i} (1 - \pi_i)^{1-y_i}, i = 1, 2, \dots, n \quad (2)$$

The likelihood function can be expressed as

$$L(\beta, y_i) = \prod_{i=1}^n f(y_i) = \prod_{i=1}^n \pi_i^{y_i} (1 - \pi_i)^{1-y_i} \quad (3)$$

Then, the log-likelihood function can be written as follows:

$$\ell(\beta, y_i) = \sum_{i=1}^n \{y_i \log \pi(x_i) + (1 - y_i) \log(1 - \pi(x_i))\} \quad (4)$$

The logistic regression is a highly discriminative classifier. However, when applying logistic regression to high-dimensional data, the predictions for the regression coefficients are inaccurate since the design matrix is singular. Aside from that, overfitting occurs in high-dimensional datasets, such as gene data sets, because the number of features (genes) is greater than the number of observations. In addition, the estimates produced by it may be negatively affected by multicollinearity [15].

It is possible that additional features may potentially create noise and degrade the classification accuracy from a statistical perspective. Because of this, researchers often seek to use feature selection techniques that may eliminate irrelevant and redundant information in order to enhance classification accuracy when developing classification models. Furthermore, one of the strategies applicable in classification in the case of high dimensionality is the penalized logistic regression technique, which is utilized to eliminate the issue of high dimensionality while simultaneously improving the accuracy of classification [16]. Despite the fact that penalization techniques are frequently employed in practice in the situation of high-dimensionality, Doerken et al. [17] have shown that these approaches can be effectively applied to data with few dimensions.

When a positive penalty component is introduced to the log-likelihood function, it is possible to drive certain coefficients to zero, resulting in a sparse solution. This is called "penalized logistic regression" (PLR) or "regularized logistic regression" (RLR) because "Regularization" is another term for this procedure. Thus, if there are too many features in the logistic model, a penalty term is included in its equation using the PLR technique. Penalizing the coefficient in this manner will decrease the coefficient values to zero. This means that the less significant features are almost equivalent to zero or precisely zero.

The regularized log-likelihood is written as

$$\text{RLR} = -\ell(\beta, y_i) + \lambda F(\beta), \quad (5)$$

where $\ell(\beta, y_i)$ indicates the log-likelihood as (4), $F(\beta)$ indicates a regularization term, and $\lambda > 0$ is a control parameter. Then the RLR of (5) is reduced with respect to the control parameter $\lambda > 0$ to get the estimates of the coefficients. This punishment reduces the variances of the estimates and imposes them to be biased, leading to enhanced prediction accuracy [18]. In classification and feature selection applications, these penalizing (regularizing) techniques belong to the family of embedded feature selection approaches that are often utilized [19].

Without loss of generality, it is assumed that the features are standardized, $(n^{-1}) \sum_{i=1}^n x_{ij}^2 = 1, \forall j \in \{1, 2, \dots, p\}$ and, $\sum_{i=1}^n x_{ij} = 0$ and the outcome variable is centered, $\sum_{i=1}^n y_i = 0$. As a result, the intercept β_0 is not penalized. The estimation of β is done by utilizing LASSO as follows.

$$\hat{\beta}_{LASSO} = \arg \min_{\beta} \left[-\sum_{i=1}^n \{y_i \log \pi(x_i) + (1 - y_i) \log(1 - \pi(x_i))\} + \lambda_1 \sum_{j=1}^p |\beta_j| \right] \quad (6)$$

where λ is the control parameter. When $\lambda = 0$, (6) is minimized to the MLE estimator. As $\lambda \rightarrow \infty$, the penalization imposes all features to be zero.

The elastic net is another effective penalized technique that is utilized in the process of feature selection. It was proposed by Zou and Hastie [9] in order to address the first and second shortcomings of LASSO, respectively. The elastic net is a combination of L_2 -norm and L_1 -norm that is used to deal with the situation of highly correlated features as well as feature selection all at the same time. The RLR with the elastic net can be expressed as follows:

$$\hat{\beta}_{Elastic} = \arg \min_{\beta} \left[-\sum_{i=1}^n \{y_i \log \pi(x_i) + (1 - y_i) \log(1 - \pi(x_i))\} + \lambda_1 \sum_{j=1}^p |\beta_j| + \lambda_2 \sum_{j=1}^p \beta_j^2 \right] \quad (7)$$

It is easy to observe from (7) that the elastic net estimator relies on two control parameters λ_1 and λ_2 whose possible values are non-negative. Equation (7) provides us with a solution to a penalized logistic regression problem.

The adaptive LASSO (ALASSO) method was originally proposed by Zou [10] to tackle the third LASSO's drawback by substituting the L_1 -regularization with a re-weighted version. In other words, Zou [10] re-weighted the L_1 -regularization coefficients. For weighting, Ridge, LASSO, or other shrinkage methods

may be used. LASSO is employed as an initial estimator for the coefficients in this work, and the LASSO obtained from a first stage is used as such. In penalized logistic models using ALASSO, the ALASSO penalized logistic model is written as

$$\hat{\beta}_{LASSO} = \arg \min_{\beta} \left[\sum_{i=1}^n \{y_i \log \pi(x_i) + (1 - y_i) \log(1 - \pi(x_i))\} + \lambda \sum_{j=1}^p \frac{|\beta_j|}{\left(|\hat{\beta}_j^{initial}\right)^{\gamma}} \right], \quad (8)$$

where $\lambda, \gamma \geq 0$ and $\hat{\beta}_j^{initial}$ is an initial estimate for each β_j estimated utilizing the LASSO method. Here, we set $\gamma = 1$, for simplicity.

Other penalized regression techniques that are similar to the elastic net method and have the capacity to accomplish grouping effect have been suggested, such as adaptive elastic net methods [11], [12], in which the authors provided two adaptive elastic net estimators. They included the adaptive weight into the L_1 -regularization when they were using the elastic net. However, in terms of adaptive weights, two adaptive elastic net (AElastic) methods vary from one another. Using the elastic net estimator, Zou and Zhang [11] build the adaptive weight. On the other hand, Ghosh [12] constructs the adaptive weight using the least squares estimator. For fixed λ_2 , the regularized logistic regression using AElastic of β can be expressed as follows:

$$\hat{\beta}_{AElastic} = \arg \min_{\beta} \left[-\sum_{i=1}^n \{y_i \log \pi(x_i) + (1 - y_i) \log(1 - \pi(x_i))\} + \lambda_1 \sum_{j=1}^p w_j |\beta_j| + \lambda_2 \sum_{j=1}^p \beta_j^2 \right] \quad (9)$$

where $w_j = (|\hat{\beta}_j|)^{-\gamma}$, $j = 1, 2, \dots, p$ is the adaptive weight generated by the initial estimator $\hat{\beta}$ for some positive constant γ . The coordinate descent method is capable of providing a reliable solution to (6)-(9) [18].

3. The Proposed Method

It has been shown that the elastic net technique performs well when the correlations between each pair of variables are extremely strong. However, El Anbari and Mkhadri [20] observed that the reliability elastic net technique lowers when the absolute correlation between features is not high ($|r| < 0.95$). Elastic net has another issue in that it ignores the correlation structure of features [21]. Zou and Zhang [11] also showed that the elastic net does not possess the oracle property and that the grouping

effect issue continues to be a concern. These issues that exist with the elastic net may be addressed by using the adaptive elastic net, which was proposed by Ghosh [12] and Zou and Zhang [11]. The adaptive elastic net combines the L_2 -regularization with the adaptive LASSO.

It is critical when using an adaptive elastic net that the initial weight is selected correctly. This is done so that features may be more accurately selected while also ensuring that classifiers are accurate. The regularized logistic regression with AElastic net (RLRAEN) is our approach that uses the adjusted variances of features as initial weights within the L_1 -regularization of the elastic net model for each feature, as shown [22].

The p -dimensional vector of features can be expressed as follows:

$$w_j = \frac{1}{|\sigma_{wj}^2|}, \quad j = 1, 2, \dots, p, \quad (10)$$

where σ_{wj}^2 is the adjusted variance of a feature j that is defined as

$$\sigma_{wj}^2 = w_{1j} \cdot \sigma_{1j}^2 + w_{2j} \cdot \sigma_{2j}^2, \quad (11)$$

where σ_{kj}^2 denotes the variance of feature j in class k , w_k is the weight of class k or prior probability. In this research, $k = 2$, and $w_1 = w_2 = 0.5$. The used weight in this research provides the feature with a low value of adjusted variance with a relatively large weight. On the other hand, it provides the feature with a high value of adjusted variance with a small weight. In this case, the L_1 -regularization can suppress inconsistency in feature selection. After ensuring that each feature has a suitable weight, the RLRAEN may use it to find related features efficiently and correctly. The RLRAEN implementation algorithm is stated in the following Algorithm. The *coordinate descent* technique may be used to well solve the RLRAEN.

Algorithm: The computation of RLRAEN

Split each descriptor X_j based on the value of Y into two classes X_{1j} and X_{2j} .

Find the variance of X_{1j} and X_{2j}

Compute $\sigma_{wj}^2 = (0.5)(\sigma_{1j}^2) + (0.5)(\sigma_{2j}^2)$

Find $w_j, j = 1, 2, \dots, p$

Define $\tilde{X}_i = W_j X_i$

Solve the RLRAEN

$$\hat{\beta}_{RLRAEN} = \arg \min_{\beta} \left[-\sum_{i=1}^n \{y_i \log \pi(x_i) + (1 - y_i) \log(1 - \pi(x_i))\} + \lambda_1 \sum_{j=1}^p w_j |\beta_j| + \lambda_2 \sum_{j=1}^p \beta_j^2 \right] \quad (12)$$

4. Experimental Study

4.1. Data Description

The proposed method employed in this study has been applied to the publicly available breast cancer Wisconsin (Diagnostic) medical dataset (WDBC). This dataset was obtained from the UCI Machine Learning dataset repository [23]. The WDBC was created by Dr. William H. Wolberg. It includes the records of 569 cases, 357 of which represent benign breast cancer and 212 cases represent malignant breast cancer. In addition to the ID number of each record and the diagnosis of breast cancer (benign and malignant), the WDBC consists of 30 real-valued features whose values were measured from a digital image of a breast mass using a special kind of needle aspirate. The values of these features represent the characteristics of the cell nuclei that appear in the image. Ten of the features are measurements of Radius, Texture, Perimeter, Area, Smoothness, Compactness, Concavity, Concave points, Symmetry, and Fractal dimension that were estimated for each cell nucleus [24], [25].

4.2. Performance Evaluation

In this study, the performance of RLRAEN in terms of prediction is assessed. Following that, a comparison with various sparse techniques is carried out. Three performance measures are included in the comparison, and they are assessed on the above dataset. These measures are classification accuracy (CA), sensitivity (SEN), and specificity (SPE) [26], [27].

$$CA = \frac{TN + TP}{FP + TP + TN + FN} \times 100\% \quad (13)$$

$$SEN = \frac{TP}{FN + TP} \times 100\% \quad (14)$$

$$SPE = \frac{TN}{TN + FP} \times 100\% \quad (15)$$

TP and *FP* indicate the number of true and false positives, respectively, while *TN* and *FN* denote the number of true and false negatives, respectively. The better the classification performance is, the higher the values of the utilized assessment criteria are. The paired t-test is used to ensure that the promised improvements have a high probability of being correct and consistent.

4.3. Experimental Setting

The proposed technique (RLRAEN) employed in this study demonstrated its effectiveness through comparative experiments with three different techniques (LASSO, elastic net, and adaptive elastic net). These techniques, along with our method, are applied to the dataset presented above. Then, cross-validation (CV) is

performed on each dataset by randomly partitioning it into two subsets: the training subset, which contains 70% of the data, and the testing subset, which contains 30% of the data. 10-fold cross-validation was performed using the training subset to obtain the best values of λ_1 and λ_2 . The experiment was performed a hundred times, and the average of the results was used as the final value. All the tuning parameters have a range of [0,100]. All the techniques used were implemented in the programming language R utilizing the "glmnet" package.

4.4. Experimental Results

To evaluate the proposed method (RLRAEN) employed in this study, we compare it with LASSO, elastic net, and AElastic net. All of these methods were applied to the WDBC dataset. The average number of features selected by each method (# features), classification accuracy, sensitivity, and specificity were calculated for both the training and testing subsets of the WDBC dataset, as is shown in Tables 1 and 2. The standard deviation for the relevant value is given in parentheses.

Table 1. Criteria (averaged over 100 times) for the training subset.

Methods	# Features	% CA	% SEN	% SPE
Lasso	14	97.13 (0.004)	97.34 (0.005)	97.19 (0.006)
AElastic	23	98.02 (0.004)	97.72 (0.004)	98.63 (0.009)
Elastic	22	97.86 (0.004)	97.73 (0.005)	98.46 (0.010)
Proposed	25	98.90 (0.005)	98.81 (0.005)	99.39 (0.004)

Table 2. Criteria (averaged over 100 times) for the testing subset.

Methods	% CA	% SEN	% SPE
Lasso	89.10 (0.010)	92.40 (0.006)	90.94 (0.006)
AElastic	91.37 (0.006)	93.13 (0.005)	90.91 (0.004)
Elastic	92.07 (0.006)	93.57 (0.007)	91.70 (0.005)
Proposed	95.94 (0.004)	94.97 (0.008)	92.98 (0.010)

At first sight, it can be seen that RLRAEN had the greatest average number of chosen features among the other three techniques when selecting 25 features on average. While LASSO, elastic net, and adaptive elastic net picked 14, 22, and 23 features on average, respectively. Furthermore, RLRAEN achieved CA, SEN, and SPE better than LASSO, elastic net, and AElastic net on average for both the training and testing sets. For instance, the training (testing) CA of RLRAEN was 98.90 (95.94), which is greater than 97.86 (92.07) for the elastic net, 97.13 (89.10) for LASSO, and 98.02 (91.37) for AElastic net.

To make sure that the evaluation is consistent, the significance of the above results was tested by means of the paired t-test. The null hypothesis is that the proposed method's "mean of average accuracy equals mean of

average accuracy of another method". The alternative hypothesis is that "the mean of average accuracy of proposed method does not equal the mean of average accuracy of another method". The column labeled "improvement" provides the relative improvement achieved by the proposed method (compared to the other methods) in mean of average accuracy. The summary of the results is presented in Table 3, which shows that RLRAEN significantly performs better than the rest of the methods. The outcomes of the study show that the proposed method, RLRAEN, outperforms some of the contemporary classifiers.

Table 3. Significant test results of paired t -test for the training and testing set.

Methods	Training set: Average accuracy		Testing set: Average accuracy	
	Improvement	p -value	Improvement	p -value
Lasso	1.86%	0.002(*)	7.68%	0.0002(*)
AElastic	0.94%	0.002(*)	5.00%	0.0001(*)
Elastic net	1.10%	0.005(*)	4.20%	0.0016(*)

(*) significant at $\alpha = 0.05$

In general, when compared to the other competitor techniques, the classification process of the proposed adaptive regularized technique offers the best overall classification process in terms of classification accuracy, sensitivity, and specificity. This indicates that RLRAEN takes into consideration the relative importance of the features.

5. Conclusions

In this study, the findings obtained by applying the RLRAEN method to the Breast Cancer dataset were compared with the findings obtained by applying the other three techniques (LASSO, elastic net, and adaptive elastic net) to the same dataset. As a result, it could be concluded that the proposed method achieved better and more efficient results for classification and feature selection than the other methods. Therefore, RLRAEN seems to be an appropriate feature selection and classification technique, and thus it may be used in other cancer-related datasets.

Acknowledgements

"We would like to express our deepest appreciation to the Taif University in Saudi Arabia for financial sponsorship".

REFERENCES

- [1] S. H. B. Hashemi, S. Karimi, and H. Mahboobi, "Lifestyle changes for prevention of breast cancer," *Electron. Physician*, vol. 6, no. 3, pp. 894–905, 2014, doi: 10.14661/2014.894-905.
- [2] V. J. Kadam, S. M. Jadhav, and K. Vijayakumar, "Breast Cancer Diagnosis Using Feature Ensemble Learning Based on Stacked Sparse Autoencoders and Softmax Regression," *J. Med. Syst.*, vol. 43, no. 8, p. 263, Aug. 2019, doi: 10.1007/s10916-019-1397-z.
- [3] W. Min, J. Liu, and S. Zhang, "Network-Regularized Sparse Logistic Regression Models for Clinical Risk Prediction and Biomarker Discovery," *IEEE/ACM Trans. Comput. Biol. Bioinforma.*, vol. 15, no. 3, pp. 944–953, May 2018, doi: 10.1109/TCBB.2016.2640303.
- [4] M. Dashtban, M. Balafar, and P. Suravajhala, "Gene selection for tumor classification using a novel bio-inspired multi-objective approach," *Genomics*, vol. 110, no. 1, pp. 10–17, Jan. 2018, doi: 10.1016/j.ygeno.2017.07.010.
- [5] S. Nakariyakul, "A hybrid gene selection algorithm based on interaction information for microarray-based cancer classification," *PLoS One*, vol. 14, no. 2, pp. 1–17, 2019, doi: 10.1371/journal.pone.0212333.
- [6] Y.-H. Yu, W. Wei, and J.-L. Liu, "Diagnostic value of fine-needle aspiration biopsy for breast mass: a systematic review and meta-analysis," *BMC Cancer*, vol. 12, no. 1, p. 41, Dec. 2012, doi: 10.1186/1471-2407-12-41.
- [7] R. Tibshirani, "Regression Shrinkage and Selection Via the Lasso," *J. R. Stat. Soc. Ser. B*, vol. 58, no. 1, pp. 267–288, Jan. 1996, doi: 10.1111/j.2517-6161.1996.tb02080.x.
- [8] J. Fan and R. Li, "Variable Selection via Nonconcave Penalized Likelihood and its Oracle Properties," *J. Am. Stat. Assoc.*, vol. 96, no. 456, pp. 1348–1360, Dec. 2001, doi: 10.1198/016214501753382273.
- [9] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," *J. R. Stat. Soc. Ser. B (Statistical Methodol.)*, vol. 67, no. 2, pp. 301–320, Apr. 2005, doi: 10.1111/j.1467-9868.2005.00503.x.
- [10] H. Zou, "The Adaptive Lasso and Its Oracle Properties," *J. Am. Stat. Assoc.*, vol. 101, no. 476, pp. 1418–1429, Dec. 2006, doi: 10.1198/016214506000000735.
- [11] H. Zou and H. H. Zhang, "On the adaptive elastic-net with a diverging number of parameters," *Ann. Stat.*, vol. 37, no. 4, pp. 1733–1751, Aug. 2009, doi: 10.1214/08-AOS625.
- [12] S. Ghosh, "On the grouped selection and model complexity of the adaptive elastic net," *Stat. Comput.*, vol. 21, no. 3, pp. 451–462, Jul. 2011, doi: 10.1007/s11222-010-9181-4.
- [13] S. Wang, B. Nan, S. Rosset, and J. Zhu, "Random lasso," *Ann. Appl. Stat.*, vol. 5, no. 1, pp. 468–485, Mar. 2011, doi: 10.1214/10-AOAS377.
- [14] Z. Y. Algamil, "Classification of gene expression autism data based on adaptive penalized logistic regression," *Electron. J. Appl. Stat. Anal.*, vol. 10, no. 2, pp. 561–571, 2017, doi: 10.1285/i20705948v10n2p561.

- [15] Z. Y. Algama, M. H. Lee, A. M. Al-Fakih, and M. Aziz, "High-dimensional QSAR classification model for anti-hepatitis C virus activity of thiourea derivatives based on the sparse logistic regression model with a bridge penalty," *J. Chemom.*, vol. 31, no. 6, p. e2889, Jun. 2017, doi: 10.1002/cem.2889.
- [16] M. El Guide, K. Jbilou, C. Koukouvinos, and A. Lappa, "Comparative study of L1 regularized logistic regression methods for variable selection," *Commun. Stat. - Simul. Comput.*, pp. 1–16, Apr. 2020, doi: 10.1080/03610918.2020.1752379.
- [17] S. Doerken, M. Avalos, E. Lagarde, and M. Schumacher, "Penalized logistic regression with low prevalence exposures beyond high dimensional settings," *PLoS One*, vol. 14, no. 5, p. e0217057, May 2019, doi: 10.1371/journal.pone.0217057.
- [18] J. Friedman, T. Hastie, and R. Tibshirani, "Regularization paths for generalized linear models via coordinate descent," *J. Stat. Softw.*, vol. 33, no. 1, pp. 1–22, 2010, doi: <https://www.ncbi.nlm.nih.gov/pubmed/20808728>.
- [19] Y. Liang *et al.*, "Sparse logistic regression with a L1/2 penalty for gene selection in cancer classification," *BMC Bioinformatics*, vol. 14, no. 1, p. 198, Dec. 2013, doi: 10.1186/1471-2105-14-198.
- [20] M. El Anbari and A. Mkhadri, "The Adaptive Gril Estimator with a Diverging Number of Parameters," *Commun. Stat. - Theory Methods*, vol. 42, no. 14, pp. 2634–2660, Jul. 2013, doi: 10.1080/03610926.2011.615438.
- [21] P. Bühlmann, P. Rütimann, S. van de Geer, and C.-H. Zhang, "Correlated variables in regression: Clustering and sparse estimation," *J. Stat. Plan. Inference*, vol. 143, no. 11, pp. 1835–1858, Nov. 2013, doi: 10.1016/j.jspi.2013.05.019.
- [22] H. Peng, Y. Fu, J. Liu, X. Fang, and C. Jiang, "Optimal gene subset selection using the modified SFFS algorithm for tumor classification," *Neural Comput. Appl.*, vol. 23, no. 6, pp. 1531–1538, Nov. 2013, doi: 10.1007/s00521-012-1148-2.
- [23] M. Lichman, "UCI Machine Learning Repository [<http://archive.ics.uci.edu/ml>].," *UCI Machine Learning Repository*, 2013. .
- [24] W. H. Wolberg, W. N. Street, and O. L. Mangasarian, "Machine learning techniques to diagnose breast cancer from image-processed nuclear features of fine needle aspirates," *Cancer Lett.*, vol. 77, no. 2–3, pp. 163–171, Mar. 1994, doi: 10.1016/0304-3835(94)90099-X.
- [25] M. C. Ferris and O. L. Mangasarian, "Breast Cancer Diagnosis via Linear Programming," *IEEE Comput. Sci. Eng.*, vol. 2, no. 3, p. 70, 1995, doi: 10.1109/MCSE.1995.414885.
- [26] A. M. Alharthi, M. H. Lee, and Z. Y. Algama, "Gene selection and classification of microarray gene expression data based on a new adaptive L1-norm elastic net penalty," *Informatics Med. Unlocked*, vol. 24, no. April, p. 100622, 2021, doi: 10.1016/j.imu.2021.100622.
- [27] A. Tharwat, "Classification assessment methods," *Appl. Comput. Informatics*, vol. 17, no. 1, pp. 168–192, Jan. 2021, doi: 10.1016/j.aci.2018.08.003.