Journal of Applied Mathematics & Bioinformatics, vol.7, no.1, 2017, 49-69
ISSN: 1792-6602 (print), 1792-6939 (online)
Scienpress Ltd, 2017

Classification of Breast Cancer Histopathology Images based on Adaptive Sparse Support Vector Machine

Mohammed Abdulrazaq Kahya[1], Waleed Al-Hayani[2] and Zakariya Yahya Algamal[3]

Abstract

Feature extraction and classification of the histopathological image plays a significant role in prediction and diagnosis of diseases, such as breast cancer. The common issues of the features matrix are that many of features may not be relevant to their diseases. Feature selection has been proved to be an effective way to improve the result of many classification methods. In this paper, an adaptive sparse support vector is proposed, with the aim of identification features, by combining the support vector machine with the weighted L1-norm. Experimental results based on a publicly recent breast cancer histopathological image datasets show that the proposed method significantly outperforms three competitor methods in terms of overall classification accuracy and the number of selected

[1] Department of Computer science, University of Mosul, Mosul, Iraq.
E-mail: mohammed.a.kahya@gmail.com:
[2] Department of Mathematics, University of Mosul, Mosul, Iraq.
E-mail: waleedalhayani@yahoo.es
[3] Department of Statistics and Informatics, University of Mosul, Mosul, Iraq.
E-mail: zakariya.algamal@gmail.com (Corresponding author)

Article Info: *Received* : February 12, 2017. *Revised* : March 17, 2017.
Published online : April 30, 2017.

features. Thus, the proposed method can be useful for medical image classification in the real clinical practice.

Mathematics Subject Classification: 68T05
Keywords: sparse support vector machine; lasso; Wilcoxon rank sum test; histopathology image; breast cancer; feature selection.

1 Introduction

According to world health organization (WHO), breast cancer among women is one of the main causes of cancer deaths in world [1]. However, the early diagnosis can increase the survival rates [2]. Several methods of noninvasive imaging namely mammograms (X-rays), magnetic resonance imaging (MRI), and ultrasonography [3, 4] are available. Recently, histopathological image analysis has become a noteworthy research problem in medical imaging [5]. Breast cancer diagnosis from a histopathology image is considered as a gold standard in diagnosing allowing to narrow borderline diagnosis issued from standard imaging methods [6].

The machine learning methods have been utilized to increase the diagnostic accuracy of women breast cancer by embedded into a computer-aided system [7]. In breast cancer classification, the taxonomy of normal and abnormal patterns of the cells is one of the most important and significant processes during the cancer diagnosis and drug discovery [8, 9]. It can help to improve the health care of patients, and, therefore, the high prediction of cancer has great value in the treatment or the therapy [10, 11].

Support vector machine is a widely-used classification method in different classification areas, especially in breast cancer classification [12]. As the number of the image features increases, the training time of applying support vector

machine increases and also its computational complexity increases [13, 14]. Unfortunately, support vector machine cannot automatically handle feature selection although it has been proven advantageous in handling binary classification [15-21].

Sparse methods are very effective embedded gene selection methods, which connected with many popular classification methods including support vector machine logistic regression, and linear discriminate analysis [22-30]. In recent years, sparse support vector machine as among all the classification methods, those based on sparseness, received much attention. It combines the standard support vector machine with a penalty to perform feature selection and classification simultaneously. With deferent penalties, several sparse support vector machine can be applied, among which are, L1-norm, which is called the least absolute shrinkage and selection operator (lasso) [31], smoothly clipped absolute deviation (SCAD) [32], elastic net [33], and adaptive L1-norm [34]. Unquestionably, L1-norm is considered to be one of the most popular procedures in the class of sparse methods. Nonetheless, L1-norm applies the same amount of the sparseness to all features, resulting in inconsistent feature selection [34-36].

To increase the power of informative feature selection, in the present study, an efficient feature selection and classification of breast cancer histopathology images, which is based on the idea of sparse support vector machine combined with Wilcoxon rank sum test, is proposed. More specifically, Wilcoxon rank sum test is employed to weight each feature. On the other hand, a sparse support vector machine with adaptive L1-norm is utilized, where each significant feature will be assigned a weight depending on the Wilcoxon rank sum test value. This weight will reflect the importance amount of each feature. Experimentally, comprehensive comparisons between our proposed method and other competitor methods are performed depending on the BreaKHis database, which contains microscopic biopsy images of benign and malignant breast tumors.

The rest of this paper is organized as follows. Section 2 explains the

preliminaries of sparse support vector machine. The Discrete wavelet transform information is explained in section 3. The proposed method with its related procedures is described in Section 4. Section 5 introduces the information of the experimental study. The experimental results are presented in Section 6. Finally, Section 7 draws general conclusions.

2 Sparse Support Vector Machine

The support vector machine (SVM), which originally proposed by Vapnik [37], is a well-known and a powerful classification method in the literature because of its strong mathematical background and excellent generalization performance. The binary classification using SVM has often been adopted in the cancer classification research because of its capability of handling nonlinear classification and high-dimensional data [8]. However, SVM itself cannot eliminate the noisy and irrelevant features [15, 17-21, 38].

Feature selection is an important tool for classifying the breast cancer. In this situation, sparse support vector machine (SSVM), which is considered as one of the embedded methods, is of more interest for researchers than the SVM because it can perform feature selection and classification simultaneously. An important SSVM is with L_1-norm (lasso) [21].

Features matrix can be described mathematically as a matrix $X = (x_{ij})_{n\times d}$, where each column represents a feature and each row represents a sample (tissue) for tumor diagnosis. The numerical value of x_{ij} denotes the value of a specific features $j\ (j = 1,...,d)$ in a specific sample $i\ (i = 1,...,n)$. Given a training dataset $\{(\mathbf{x}_i, y_i)\}_{i=1}^{n}$, where $\mathbf{x}_i = (x_{\mathrm{i},1}, x_{\mathrm{i},2},...,x_{\mathrm{i,d}})$ represents a vector of the i^{th} feature, and $y_i \in \{-1,+1\}$ for $i = 1,...,n$, where $y_i = +1$ indicates the i^{th} sample is in class 1 (e.g., has cancer) and $y_i = -1$ indicates the i^{th} sample is in

class 2 (e.g., does not have cancer). Generally, the objective is to classify the new sample and identify the relevant feature with high classification accuracy.

The classical SVM solves the optimization problem by minimizing

$$\frac{1}{n}\sum_{i=1}^{n}\left[1-y_i\left(b+\mathbf{z}\,h(\mathbf{x}_i)\right)\right]_+ + \lambda \|\mathbf{z}\|_2^2 \tag{1}$$

where $\left[1-y_i\left(b+\mathbf{z}\,h(\mathbf{x}_i)\right)\right]_+$ is the convex hinge loss, the scalar b is denoted as the bias, $\|\mathbf{z}\|_2^2$ is the L_2-norm, and $\lambda > 0$ is the tuning parameter controlling the trade-off between minimizing the hyper-plane coefficients and the classification error. Equation (2) is a convex optimization problem and can be solved by the method of Lagrange multipliers [15]. The optimization solution can provide a unique solution for hyperplane parameters $\mathbf{z}$ and b.

Although SVM is a widely-used classification method in different classification areas, it cannot perform feature selection because of using L_2-norm. This can be a downside when there are many irrelevant features [15, 17, 20, 21, 39]. To overcome this limitation, those methods for simultaneous feature selection and classification are more preferable to achieve better classification accuracy with less important features [17].

For the purpose of feature selection, several variants of penalties are adopting with SVM. Bradley and Mangasarian [40] and Zhu, Rosset, Hastie and Tibshirani [21] proposed using L_1-norm instead of L_2-norm of Eq. (1) to perform variable selection and binary classification. Ikeda and Murata [41], Liu, Helen Zhang, Park and Ahn [17], and Liu, Lin and Tan [18] proposed L_q-norm with $q < 1$. Furthermore, Zhang, Ahn, Lin and Park [20] proposed the smoothly clipped absolute deviation (SCAD) penalty of Fan and Li [32] with SVM. In addition, Wang, Zhu and Zou [38] proposed a hybrid huberized SVM by using the elastic net penalty. Becker, Toedt, Lichter and Benner [15] proposed a combination of ridge and SCAD with SVM.

Because of the singularity of the L_1-norm, SVM with L_1-norm can automatically select variables by shrinking the hyper-plane coefficients to zero [15,

38]. In addition, SCAD has the same behavior as L_1-norm [15]. The SSVM with L1-norm (SSVM-lasso) and the SSVM with SCAD (SSVM-SCAD) are, respectively, defined as

$$\frac{1}{n}\sum_{i=1}^{n}\left[1-y_i\left(b+\mathbf{z}\,h(\mathbf{x}_i)\right)\right]_+ + \lambda|\mathbf{z}| \tag{2}$$

$$\frac{1}{n}\sum_{i=1}^{n}\left[1-y_i\left(b+\mathbf{z}\,h(\mathbf{x}_i)\right)\right]_+ + pen(\mathbf{z}) \tag{3}$$

where

$$pen(\mathbf{z})=\begin{cases}\lambda|z_j| & if\ \ |z_j|\le\lambda,\\ -\dfrac{|z_j|^2-2a\lambda|z_j|+\lambda^2}{2(a-1)} & if\ \ \lambda<|z_j|\le a\lambda,\\ \dfrac{(a+1)\lambda^2}{2} & if\ \ |z_j|>a\lambda,\end{cases} \tag{4}$$

where z_j, $j=1,2,...,p$ are the hyper-plane coefficients, $a=3.7$ as suggested by Fan and Li [32], and $\lambda>0$ is the tuning parameter.

3 Discrete Wavelet Transform

The wavelet series is just a sampled version of continuous wavelet transform (CWT) and its computation may consume a significant amount of time and resources, depending on the resolution required. The discrete wavelet transform (DWT), which is based on sub-band coding is found to yield a fast computation of wavelet transform. It is easy to implement and reduces the computation time and resources required [42].

A two-dimensional scaling function, $\varphi(x,y)$, and three two-dimensional wavelet $\psi^H(x,y)$, $\psi^V(x,y)$ and $\psi^D(x,y)$ are critical elements for wavelet transforms in two dimensions [43]. Given separable 2-D scaling and wavelet functions, 2-D DWT can be defined as: First, we define the scaled and translated

or shifted basis functions which are defined as follows [43]:

$$\varphi_{j,m,n}(x,y) = 2^{j/2}\varphi(2^j x - m, 2^j y - n) \tag{5}$$

$$\psi^i{}_{j,m,n}(x,y) = 2^{j/2}\psi^i\left(2^j x - m, 2^j y - n\right) \qquad i = \{H, V, D\} \tag{6}$$

where i is the directional wavelet index. Therefore, 2-D DWT of an image $f(x,y)$ of size $M \times N$ is given by [11]:

$$W\varphi(j_0,m,n) = \frac{1}{\sqrt{M\,N}}\sum_{x=0}^{M-1}\sum_{y=0}^{N-1} f(x,y)\varphi_{j_0,m,n}(x,y) \tag{7}$$

$$W\psi^i(j,m,n) = \frac{1}{\sqrt{M\,N}}\sum_{x=0}^{M-1}\sum_{y=0}^{N-1} f(x,y)\psi^i{}_{j,m,n}(x,y) \tag{8}$$

where j_0 is an arbitrary starting scale, $W\varphi(j_0,m,n)$ is the approximation coefficients for $f(x,y)$ at scale j_0, $W\psi^i(j,m,n)$ is the horizontal, vertical and diagonal details coefficients at scales $j \geq j_0$, and $M = N = 2^j$ for $j_0 = 0,1,\ldots, j-1$, $m,n = 0,1,\ldots,2^j - 1$. Then the two-dimensional DWT can be implemented using digital filters and downsamplers $2\downarrow$.

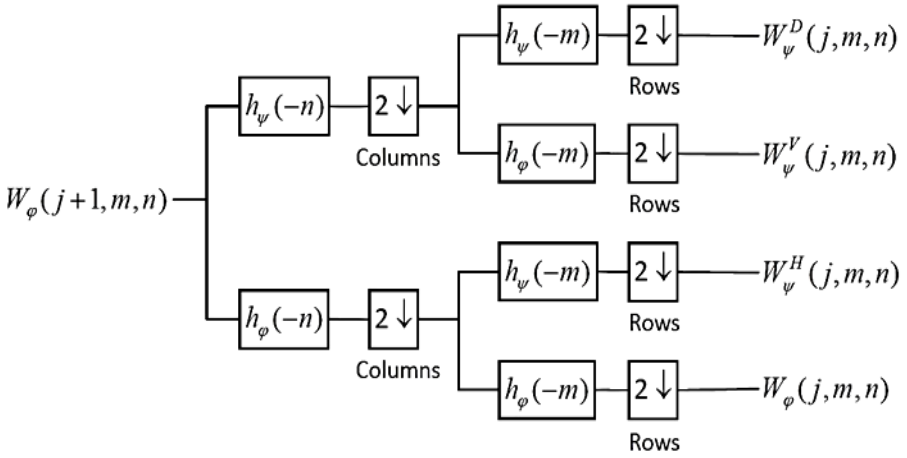

Figure 1: The analysis filter bank of the two-dimensional FWT

The block diagram in Figure 1 shows the process of taking the one-dimensional

FWT of the rows of $f(x,y)$ and the subsequent one-dimensional FWT of the resulting columns. Three sets of detail coefficients including the horizontal, vertical, and diagonal details are produced.

Depending on the discrete wavelet transform, specifically, Haar discrete wavelet transform based on the 7^{th} level decomposition, was employed to extract the features of the breast cancer histopathology images. At first level of decomposition, breast cancer histopathology images were divided into four equal size sub-images, namely LL1 (approximation coefficients), LH1 (horizontal coefficients), HL1 (vertical coefficient) and HH1 (diagonal coefficient). Subsequently at the second level of decomposition LL1 (approximation coefficient) sub-image is further decomposed into four equal size sub-images LL2, LH2, HL2 and HH2. Continuously until we reach the seventh level of decomposition. In this manner 28 sub-images have been formed from every channel (red, green and blue). Thus, 28 x 3 sub-images have been established from the original image. Then nine of the traditional statistical features (Mean, Standard deviation, skewness, kurtosis, Entropy, energy, root mean square, Mean absolute deviation, Median Absolute Deviation) are calculated. As a result, 756 features are extracted for each histopathology image of the breast cancer.

4 The Proposed Method

In the context of breast cancer classification, the goal of feature selection is to improve classification performance. High dimensionality of features can negatively influence the classification performance of a classifier by increasing the risk of overfitting and lengthening the computational time. Therefore, removing irrelevant and noisy features from the original features matrix is essential for applying classification methods.

It is worthwhile to highlight that our contribution of this paper comes from

the following issue. Although SSVM with L_1-norm can be applied directly to the breast cancer classification, this method may select irrelevant features because L_1-norm has the inconsistent property in feature selection. In other words, the estimates of the SSVM with L_1-norm can be biased for large hyper-plane coefficients because larger coefficients will take larger penalties. Compared with L_1-norm, SSVM with SCAD generally suffer from non-convexity although SSVM with SCAD proved its consistency in feature selection.

Consequently, efficient feature selection is proposed. It is based on the idea of SSVM with L_1-norm combined with Wilcoxon rank sum test. More specifically, Wilcoxon rank sum test is employed to weight each feature. On the other hand, the SSVM with adaptive L_1-norm is utilized, where each significant feature will be assigned a weight depending on the Wilcoxon rank sum test value. This weight will reflect the importance amount of each feature.

4.1 Weight Calculation

In practice, feature matrix contains irrelevant or noisy features leading to low performance with less classification accuracy. As a consequence, analyzing feature in terms of their importance has become a necessary task. To determine the weight for each feature, the Wilcoxon rank sum test [44] is utilized as

$$s(j) = \sum_{i \in N_1} \sum_{k \in N_2} I((\mathbf{x}_i^{(j)} - \mathbf{x}_k^{(j)}) \leq 0), \quad j = 1, 2, ..., p \tag{9}$$

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

$$I(\cdot) = \begin{cases} 1 & \text{if } I \text{ is true} \\ 0 & \text{if } I \text{ is not true} \end{cases} \tag{10}$$

$\mathbf{x}_i^{(j)}$ is the value of the sample i in the feature j, and N_1 and N_2 are the index sets of different classes of samples. Equation (5), $s(j)$, represents the measurement of the difference between the two classes. The feature j can be

considered important when Eq. (5) 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 [44] quantify the feature significance by the following feature ranking criterion

$$q(j) = \max\{s(j), n_1 n_2 - s(j)\}. \quad (11)$$

Depending on Eq. (7), 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, Shiraishi, Imoto and Miyano [45] proposed the following weight

$$w_j = 1/[\frac{q(j)}{\sum_{j=1}^{p} q(j)} * p], \ j = 1, 2, ..., p. \quad (12)$$

According to Equation (8), 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.

4.2 Breast Cancer Classification

After assigning each feature with its related weight, the SSVM with adaptive L_1-norm is utilized to select the informative features with high classification accuracy. The detailed of the adaptive SSVM (ASSVM) computation is described in Algorithm 1. The ASSVM equation has a convex form, which ensures the existence of global maximum point and can be efficiently solved.

Algorithm 1: The computation of ASSVM

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

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

Step 3: Solve the ASSVM,

$$\frac{1}{n}\sum_{i=1}^{n}\left[1 - y_i(b + \mathbf{z}\, h(\tilde{\mathbf{x}}_i))\right]_+ + \lambda|\mathbf{z}|.$$

5 Experimental Study

5.1 Datasets Description

The dataset that has been exploited is related to The BreaKHis database, which contains microscopic biopsy images of benign and malignant breast tumors [4]. This dataset is related to the pathological anatomy and cytopathology laboratory of Parana, Brazil. This database, BreaKH, is composed of 7909 clinically representative microscopic images of breast tumor tissue images collected from 82 patients using different magnifying factors: 40X, 100X, 200X, and 400X, with 24 benign and 58 malignant samples. A summary of this database is listed in Table 1.

Table 1: Summary of the BreaKH database.

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

5.2 Performance Evaluation

In order to evaluate the predictive performance of the proposed method, two performance metrics are implemented, specifically: (1) patient classification rate

(PCR) and (2) overall classification accuracy (OCA). The PCR stands for the proportion of correctly classified benign class and malignant class within the patient. The PCR can define as:

$$\mathrm{PCR} = \frac{\mathrm{n}_{correct}}{\mathrm{n}_{total}} \times 100\% \tag{13}$$

where $\mathrm{n}_{correct}$ 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:

$$\mathrm{OCA} = \frac{\sum_{j=1}^{p} PCR_j}{\mathrm{n}_{patients}} \times 100\% \tag{14}$$

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

5.3 Experimental Setting

To demonstrate the usefulness of the proposed method, comprehensive comparative experiments with the SSVM-lasso, SSVM-SCAD, and the classical SVM are conducted. To do so, the data matrix is randomly partitioned into the training dataset and the test dataset, where 70% of the samples are selected for training dataset and the rest 30% are selected for testing dataset. For a fair comparison and for alleviating the effect of the data partition, all the used classification methods are evaluated, for their classification performance metrics using 10 folds cross validation, averaged over 10 partitioned times.

Depending on the training dataset, the tuning parameter value, λ, for each used classification method was fixed as $0 \le \lambda \le 100$. For the SCAD penalty, the constant a was set to equal 3.7 as it suggested by Fan and Li [32]. The implementations of these used methods are provided in the R-package: penalized SVM.

6 Experimental Results

6.1 Classification Performance

Table 2 summarizes, on average, the overall classification accuracy for the training and testing datasets of applying the ASSVM, SSVM-SCAD, SSVM-lasso, and the SVM. 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, ASSVM, achieves 95.37%, defeating SSVM-SCAD, SSVM-lasso, and the SVM by 4.40%, 5.50%, and 7.63%, respectively. In addition, SSVM-SCAD secondly comes with 90.97% and better than SSVM-lasso and SVM. Depending on the testing dataset, the ASSVM is better than the others in terms of overall classification accuracy because it achieved 94.97%, which is 6.37%, 7.95%, and 10.64% better than SSVM-SCAD, SSVM-lasso, and the SVM, respectively.

In the magnification 100X, based on the training dataset, the ASSVM provides enhancement over the SSVM-SCAD and the SSVM-lasso by 3.82% and 5.91%, respectively. Once again, based on the testing dataset, the proposed method beats both SSVM-SCAD and SSVM-lasso in terms of overall classification accuracy.

Looking at the magnification 200X, the overall classification performance of the proposed method is comparable with SSVM-SCAD, SSVM-lasso, and SVM performing best among them. In terms of overall classification accuracy, the OCA obtained from the proposed method was 96.28% for the training dataset and 94.54% for the testing dataset. This indicates the superiority of the proposed method as compared to SSVM-SCAD, SSVM-lasso, and SVM.

At the end, regarding the magnification 400X, the ASSVM shows a considerable dominance against SSVM-SCAD, SSVM-lasso, and SVM. It achieved the higher overall classification accuracy for both the training and testing

datasets. SSVM-SCAD and SSVM-lasso attain mediocre performance as they provide results that are inferior to AHP but better than SVM.

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 ASSVM outperformed SSVM-SCAD and SSVM-lasso. For instance, in magnification 40X, ASSVM selected 6 features compared to 17 and 22 features for the SSVM-SCAD and SSVM-lasso, respectively.

Table 2: Classification performance of the ASSVM, SSVM-SCAD, SSVM-lasso and SVM

Methods	Training dataset	Testing dataset	# selected features
	OCA	OCA	
40X			
ASSVM	95.37 (0.09)	94.97 (0.003)	6
SSVM-SCAD	90.97 (0.011)	88.60 (0.007)	17
SSVM-lasso	89.87 (0.011)	87.02 (0.007)	22
SVM	87.74 (0.013)	84.33 (0.007)	All
100X			
ASSVM	95.75 (0.008)	93.62 (0.007)	10
SSVM-SCAD	91.93 (0.008)	88.49 (0.008)	19
SSVM-lasso	89.84 (0.008)	86.15 (0.008)	24
SVM	85.60 (0.011)	81.80 (0.009)	All
200X			
ASSVM	96.28 (0.008)	94.54 (0.005)	8
SSVM-SCAD	91.02 (0.011)	89.56 (0.007)	14

SSVM-lasso	86.29 (0.011)	89.35 (0.007)	20
SVM	86.28 (0.013)	82.39 (0.008)	All
400X			
ASSVM	95.68 (0.008)	94.42 (0.005)	8
SSVM-SCAD	90.63 (0.011)	88.68 (0.007)	13
SSVM-lasso	88.59 (0.011)	84.96 (0.007)	19
SVM	84.28 (0.013)	80.91 (0.008)	All

6.2 Statistical Significance Test

For further ability confirmation of the proposed method in selecting the most relevant features with high classification performance, a pairwise comparison between the proposed method and each competitor method was utilized using Mann–Whitney U test. This test was performed depending on the area under the curve (AUC) of the training dataset. Table 3 reports the Mann–Whitney U test results at significance level $\alpha = 0.05$. As shown in Table 3, the AUC of the proposed method is statistically significant better than those of SSVM-SCAD, SSVM-lasso and SVM.

Table 3: P-values for the Mann–Whitney U test of our proposed method results with three competitor methods. (*) means that the two methods have significant differences

Pairwise comparison	p-value
ASSVM vs SSVM-SCAD	0.0017(*)
ASSVM vs SSVM-lasso	0.0011(*)
ASSVM vs SVM	0.0001(*)

7 Conclusion

This paper presents an adaptive sparse support vector machine by combining the support vector machine with the weighted L_1-norm to classify the breast cancer histopathology images. Our proposed method was experimentally tested and compared with other existing methods. The superior classification performance of the proposed method was shown through the overall classification accuracy and the Mann–Whitney U test for the AUC. Furthermore, ASSVM performs remarkably well in terms of the number of selected features as compared to SSVM-SCAD, SSVM-lasso. Consequently, the results confirm that ASSVM is a promising feature selection method for medical image classification.

Funding: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

[1] B.K. Singh, K. Verma, A.S. Thoke, Fuzzy cluster based neural network classifier for classifying breast tumors in ultrasound images, *Expert. Syst. Appl.*, **66**, (2016), 114-123.

[2] V.K. Sudarshan, M.R. Mookiah, U.R. Acharya, V. Chandran, F. Molinari, H. Fujita, K.H. Ng, Application of wavelet techniques for cancer diagnosis using ultrasound images: A Review, *Comput. Biol. Med.*, **69**, (2016), 97-111.

[3] U.R. Acharya, U. Raghavendra, H. Fujita, Y. Hagiwara, J.E. Koh, T. Jen Hong, V.K. Sudarshan, A. Vijayananthan, C.H. Yeong, A. Gudigar, K.H. Ng, Automated characterization of fatty liver disease and cirrhosis using curvelet transform and entropy features extracted from ultrasound images, *Comput. Biol. Med.*, **79**, (2016), 250-258.

[4] F.A. Spanhol, L.S. Oliveira, C. Petitjean, L. Heutte, A dataset for breast cancer histopathological image classification, *IEEE Transactions on Biomedical Engineering*, **63**, (2016), 1455-1462.

[5] T.H. Vu, H.S. Mousavi, V. Monga, G. Rao, U.K. Rao, Histopathological image classification using discriminative feature-oriented dictionary learning, *IEEE Transactions on Medical Imaging*, **35**, (2016), 738-751.

[6] Y. Zhang, B. Zhang, W. Lu, Breast cancer histological image classification with multiple features and random subspace classifier ensemble, in: T.D. Pham, L.C. Jain (Eds.) Knowledge-Based Systems in Biomedicine and Computational Life Science, Springer Berlin Heidelberg, Berlin, Heidelberg, (2013), 27-42.

[7] A.D. Belsare, M.M. Mushrif, M.A. Pangarkar, N. Meshram, Classification of breast cancer histopathology images using texture feature analysis, TENCON 2015 - 2015 IEEE Region 10 Conference, (2015), 1-5.

[8] S. Korkmaz, G. Zararsiz, D. Goksuluk, Drug/nondrug classification using Support Vector Machines with various feature selection strategies, *Comput. Methods. Programs. Biomed.*, **117**, (2014), 51-60.

[9] L. Zhang, L. Qian, C. Ding, W. Zhou, F. Li, Similarity-balanced discriminant neighbor embedding and its application to cancer classification based on gene expression data, *Comput. Biol. Med.*, **64**, (2015), 236-245.

[10] Y. Chen, L. Wang, L. Li, H. Zhang, Z. Yuan, Informative gene selection and the direct classification of tumors based on relative simplicity, *BMC Bioinformatics*, **17**, (2016), 44-57.

[11] Z. Mao, W. Cai, X. Shao, Selecting significant genes by randomization test for cancer classification using gene expression data, *J. Biomed. Inform.*, **46**, (2013), 594-601.

[12] B. Zheng, S.W. Yoon, S.S. Lam, Breast cancer diagnosis based on feature extraction using a hybrid of K-means and support vector machine algorithms, *Expert. Syst. Appl.*, **41**, (2014), 1476-1482.

[13] M. Kumar, S. Kumar Rath, Classification of microarray using MapReduce based proximal support vector machine classifier, *Knowl.-Based. Syst.*, **89**, (2015), 584-602.

[14] X. Peng, D. Xu, L. Kong, D. Chen, L1-norm loss based twin support vector machine for data recognition, *Inform. Scienc.*, **340-341**, (2016), 86-103.

[15] N. Becker, G. Toedt, P. Lichter, A. Benner, Elastic SCAD as a novel penalization method for SVM classification tasks in high-dimensional data, *BMC bioinformatics*, **12**, (2011), 138-151.

[16] J. Bi, K.P. Bennett, M. Embrechts, C. Breneman, M. Song, Dimensionality reduction via sparse support vector machines, *J. Mach. Learn. Res.*, **3**, (2003), 1229-1243.

[17] Y. Liu, H. Helen Zhang, C. Park, J. Ahn, Support vector machines with adaptive Lq penalty, *Comput. Stat. Data. Anal.*, **51**, (2007), 6380-6394.

[18] Z. Liu, S. Lin, M.T. Tan, Sparse support vector machines with Lp penalty for biomarker identification, *IEEE Trans. Comput. Bi.*, **7**, (2010), 100-107.

[19] C. Park, K.-R. Kim, R. Myung, J.-Y. Koo, Oracle properties of SCAD-penalized support vector machine, *J. Stat. Plan. Infer.*, **142**, (2012), 2257-2270.

[20] H.H. Zhang, J. Ahn, X. Lin, C. Park, Gene selection using support vector machines with non-convex penalty, *Bioinformatics*, **22**, (2006), 88-95.

[21] J. Zhu, S. Rosset, T. Hastie, R. Tibshirani, 1-norm support vector machines, *Adv. Neural. Inf. Process. Syst.*, **16**, (2004), 49-56.

[22] Y. Cui, C.H. Zheng, J. Yang, W. Sha, Sparse maximum margin discriminant analysis for feature extraction and gene selection on gene expression data, *Comput. Biol. Med.*, **43**, (2013), 933-941.

[23] Z.Y. Algamal, M.H. Lee, A new adaptive L1-norm for optimal descriptor selection of high-dimensional QSAR classification model for anti-hepatitis C virus activity of thiourea derivatives, *SAR QSAR Environ Res*, **28**, (2017), 75-90.

[24] Z.Y. Algamal, M.H. Lee, A.M. Al-Fakih, High-dimensional quantitative structure-activity relationship modeling of influenza neuraminidase a/PR/8/34 (H1N1) inhibitors based on a two-stage adaptive penalized rank regression, *J. Chemom.*, **30**, (2016), 50-57.

[25] Z.Y. Algamal, M.H. Lee, A.M. Al-Fakih, M. Aziz, High-dimensional QSAR modelling using penalized linear regression model with L1/2-norm, *SAR QSAR Environ Res*, **27**, (2016), 703-719.

[26] Z.Y. Algamal, M.H. Lee, A.M. Al-Fakih, M. Aziz, High-dimensional QSAR prediction of anticancer potency of imidazo[4,5-b]pyridine derivatives using adjusted adaptive LASSO, *J. Chemom.*, **29**, (2015), 547-556.

[27] Z.Y. Algamal, M.H. Lee, Applying penalized binary logistic regression with correlation based elastic net for variables selection, *J. Mode. Appl. Stat. Meth.*, **14**, (2015), 168-179.

[28] Z.Y. Algamal, M.H. Lee, High dimensional logistic regression model using adjusted elastic net penalty, *Pakistan J. Stat. & Oper. Rese.*, **11**, (2015), 667-676.

[29] V. Pappua, O.P. Panagopoulosb, P. Xanthopoulosb, P.M. Pardalosa, Sparse proximal support vector machines for feature selection in high dimensional datasets, *Expert. Syst. Appl.*, **42**, (2015), 9183–9191.

[30] S.K. Shevade, S.S. Keerthi, A simple and efficient algorithm for gene selection using sparse logistic regression, *Bioinformatics*, **19**, (2003), 2246-2253.

[31] R. Tibshirani, Regression shrinkage and selection via the lasso, *J. Roy. Statist. Soc. Ser. B*, **58**, (1996), 267-288.

[32] J. Fan, R. Li, Variable selection via nonconcave penalized likelihood and its oracle properties, *J. Amer. Statist. Assoc.*, **96**, (2001), 1348-1360.

[33] H. Zou, T. Hastie, Regularization and variable selection via the elastic net, *J. Roy. Statist. Soc. Ser. B*, **67**, (2005), 301-320.

[34] H. Zou, The adaptive lasso and its oracle properties, *J. Amer. Statist. Assoc.*, **101**, (2006), 1418-1429.

[35] Z.Y. Algamal, M.H. Lee, Penalized logistic regression with the adaptive LASSO for gene selection in high-dimensional cancer classification, *Expert. Syst. Appl.*, **42**, (2015), 9326–9332.

[36] Z.Y. Algamal, M.H. Lee, Regularized logistic regression with adjusted adaptive elastic net for gene selection in high dimensional cancer classification, *Comput. Biol. Med.*, **67**, (2015), 136-145.

[37] V. Vapnik, *The nature of statistical learning theory*, Springer Science & Business Media, New York, NY, USA, 1995.

[38] L. Wang, J. Zhu, H. Zou, Hybrid huberized support vector machines for microarray classification and gene selection, *Bioinformatics*, **24**, (2008), 412-419.

[39] S. Maldonado, R. Montoya, J. López, Embedded heterogeneous feature selection for conjoint analysis: A SVM approach using L1 penalty, *Appli. Intellig.*, DOI 10.1007/s10489-016-0852-5, (2016).

[40] P.S. Bradley, O.L. Mangasarian, Feature selection via concave minimization and support vector machines, ICML, (1998), 82-90.

[41] K. Ikeda, N. Murata, Geometrical properties of Nu support vector machines with different norms, *Neur. Comput.*, **17**, (2005), 2508-2529.

[42] B. Banoth, Rotation and scale invariant texture classification using log-polar wavelet energy signatures, A thesis submitted for the degree Master of Technology in Telematics and Signal Processing, National Institute of Technology Rourkela, India, 2007.

[43] R.C. Gonzalez, R.E. Woods, *Digital image processing*, Prentice Hall, New York, 2008.

[44] C. Liao, S. Li, Z. Luo, Gene selection using Wilcoxon rank sum test and support vector machine for cancer classification, in: Y. Wang, Y.-m. Cheung, H. Liu (Eds.) Computational Intelligence and Security: International

Conference, CIS 2006. Guangzhou, China, November 3-6, 2006. Revised Selected Papers, Springer Berlin Heidelberg, Berlin, Heidelberg, (2007), 57-66.

[45] H. Park, Y. Shiraishi, S. Imoto, S. Miyano, A novel adaptive penalized logistic regression for uncovering biomarker associated with anti-cancer drug sensitivity, *IEEE Trans. Comput. Bi.*, DOI 10.1109/TCBB.2016.2561937, (2016).