Research Article

Sparse Representation for Tumor Classification Based on Feature Extraction Using Latent Low-Rank Representation

Bin Gan,[1] Chun-Hou Zheng,[1,2] Jun Zhang,[2] and Hong-Qiang Wang[3]

[1] *College of Information and Communication Technology, Qufu Normal University, Rizhao 276800, China*
[2] *College of Electrical Engineering and Automation, Anhui University, Hefei 230000, China*
[3] *Intelligent Computing Lab, Institute of Intelligent Machines, Chinese Academy of Sciences, Hefei 230000, China*

Correspondence should be addressed to Chun-Hou Zheng; zhengch99@126.com and Hong-Qiang Wang; hqwang@ustc.edu

Accurate tumor classification is crucial to the proper treatment of cancer. To now, sparse representation (SR) has shown its great performance for tumor classification. This paper conceives a new SR-based method for tumor classification by using gene expression data. In the proposed method, we firstly use latent low-rank representation for extracting salient features and removing noise from the original samples data. Then we use sparse representation classifier (SRC) to build tumor classification model. The experimental results on several real-world data sets show that our method is more efficient and more effective than the previous classification methods including SVM, SRC, and LASSO.

1. Introduction

Tumor is a solid lesion caused by the abnormal growth of cells. A timely accurate treatment is very important clinically. The premise of an accurate treatment is an exact diagnosis due to the heterogeneity of cancer. That is, we need to classify them accurately before treating tumors. Current methods for classifying cancer malignancies mostly rely on a variety of morphological, clinical, or molecular variables. Despite recent progresses, there are still many uncertainties in diagnosis. The advent of DNA microarray and RNA_seq [1] makes it possible to analyze tumor samples and classify them based on gene expression profiles. Moreover, we can get the expression data of tens of thousands of genes through DNA microarray or RNA-seq simultaneously.

Many methods for molecular data classification or clustering based on gene expression data have appeared in this area [2–14]. Huang and Zheng used independent component analysis [5] to extract features; Gao and Church introduced sparse nonnegative matrix factorization for feature extraction [4]; Zheng et al. proposed metasample-based sparse representation [7], and Furey et al. used support vector machines [8] to classify the gene expression data. All these methods have achieved impressive classification performances.

Recently published sparse representation classification (SRC) is also a powerful tool for processing gene expression data. SRC method was inspired by many theories such as Basis pursuing [15], compressive sensing for signal reconstruction [16], and least absolute shrinkage. It has already been widely used in face recognition [17] and texture classification [18]. In SRC method, test samples can be only represented as a sparse linear combination of the training samples from the same class. Furthermore, an imposed l_1-regularized least square optimization is used to calculate an SR coefficient vector with only a few significant coefficients. In theory, a test sample can be well represented by only using the training samples from the same class. However, there is too much noise in gene expression data, which causes that the discriminative features are not obvious and the test samples can also be represented by some training samples from different classes. This will decrease the classification accuracy. To reduce noise [19–21] and get salient features [20] for tumor classification, in this paper, we introduce latent

low-rank representation to preprocess gene expression data. By combining it with SRC algorithm, we propose a new method for tumor classification.

Latent low-rank representation (LatLRR) is a kind of theory which can be used to extract principal and salient features from original data. LatLRR is the improved version of LRR. The two methods can be resolved by the inexact augmented Lagrange multiplier (ALM) optimization. In [19–22], LRR has been successfully used for the recovery of subspace structure, subspace segmentation, feature extraction, outlier detection, and so forth. In [23], the author introduced LRR theory for face recognition in order to remove noise and achieved an impressive result. Based on these successful applications, in this paper, we introduce LatLRR into sparse representation classifier for tumor classification. Firstly, we use LatLRR to remove noise from original data and extract salient features. Then based on the new extracted salient features, we design sparse representation classifier to classify new test samples. We referred to the proposed method as SRC-based latent low-rank representation (SRC- LatLRR).

The rest of the paper is organized as follows. Section 2 describes our proposed SRC-LatLRR method in detail. We firstly review SRC and latent low-rank representation methods in Sections 2.1 and 2.2, respectively. Then we present our method in detail in Section 2.3. Section 2.4 specifies our experimental setting. In Section 3, we evaluate our method using several publicly available gene expression data sets. Section 4 concludes the paper and outlines our future work.

The abbreviations used in this paper are summarized in the Abbreviations section.

2. Methods

2.1. Sparse Representation Classification.
Sparse representation classification is a supervised classification. Let $W \in R^{m \times n}$ denote a training sample matrix with n samples and m genes. As we know, each DNA microarray chip usually contains thousands of genes; the number of genes is much larger than tumor samples; that is, $m \gg n$.

Let c_l be the lth sample of W and the n samples are divided into k object classes. Assuming that there are n_i samples belonging to ith class and making up $W_i = [c_{i,1}, c_{i,2}, \ldots, c_{i,n_i}]$, the whole data set can be reexpressed as $W = [W_1, W_2, \ldots, W_k]$. Suppose that a new testing sample $y \in R^m$ belongs to ith class. Based on the theory of sparse representation, y would lie in the linear span of the training samples W_i; that is,

$$y = \alpha_{i,1} c_{i,1} + \alpha_{i,2} c_{i,2} + \cdots + \alpha_{i,n_i} c_{i,n_i}, \qquad (1)$$

where $\alpha_{i,j} \in R$ is a scalar and $j = 1, 2, \ldots, n_i$.

Supposing a linear representation coefficient vector $x_0 \in R^n$, y can be also rewritten as

$$y = W x_0. \qquad (2)$$

Ideally, if the training samples are sufficient and the training samples sets that belong to different class are disjoint each other, then we have

$$x_0 = [0, \ldots, 0, \alpha_{i,1}, \alpha_{i,2}, \ldots, \alpha_{i,n_i}, 0, \ldots, 0] \in R^n; \qquad (3)$$

that is, in x_0, only the entries corresponding to the same class as y are nonzero.

From the above analysis, it can be seen that we can classify the test sample y according to x_0. So the key problem is how to calculate x_0 in (2). As in [7], x_0 would be sparse if the number of object classes k is large; this is what sparse representation implies. According to the theory of compressive sensing [16, 24–26] and SR, x_0 can be achieved by solving the following l_1-minimization problem:

$$\hat{x}_1 = \arg \min \|x\|_1 \quad \text{s.t. } W x = y. \qquad (4)$$

This problem can be solved by standard linear programming methods [15]. But (4) has no exact solutions since $m \gg n$. Then a generalized version of (4) can be conceived:

$$J(x, \lambda) = \min_x \left\{ \|W x - y\|_2 + \lambda \|x\|_1 \right\}, \qquad (5)$$

where λ is a scalar regularization. This function can balance the degree of noise by using λ. In this study, we solve this function by the truncated Newton interior-point method [27].

2.2. Latent Low-Rank Representation.
Latent low-rank representation is an extension of low-rank representation. Consider an observed data matrix $X = [x_1, x_2, \ldots, x_n] \in R^{D \times n}$, where each column vector x_i is a sample, and a dictionary $A = [a_1, a_2, \ldots, a_m] \in R^{D \times m}$, where a_i is also a sample. X can be linearly represented by the dictionary. That is,

$$X = AZ, \qquad (6)$$

where $Z = [z_1, z_2, \ldots, z_n] \in R^{m \times n}$ is a coefficient matrix and each z_i is the representation of x_i. Equation (6) means that each column vector of X can be represented by a linear combination of the bases in A. In (6), the dictionary A should be overcomplete enough to represent any observed data matrix X. But meanwhile, this causes multiple feasible solutions of Z to (6). To achieve the optimal solution, low rankness criterion is introduced to (6):

$$\min_Z \text{rank}(Z), \quad \text{s.t. } X = AZ. \qquad (7)$$

Here, the optimal solution Z^* is the so-called lowest-rank representation of data X with respect to the dictionary A. Unfortunately, function (7) can not be easy to solve because of the discrete nature of the rank function. By matrix completion method [28–30], we replace solving low-rank problem with dealing with nuclear norm [31]; then problem (7) can be rerepresented as

$$\min_Z \|Z\|_*, \quad \text{s.t. } X = AZ, \qquad (8)$$

where $\|Z\|_*$ means the nuclear norm of matrix Z, that is, the sum of the singular values of matrix Z.

Strictly speaking, the dictionary A should be overcomplete and noiseless. But this kind of dictionary is difficult to get. In practice, we usually use observed data matrix X itself

as the dictionary [19, 21, 32]. Finally we have the following convex optimization problem:

$$\min_Z \|Z\|_*, \quad \text{s.t. } X = XZ. \tag{9}$$

To solve this equation, two conditions need to be met. Firstly, the data sampling X should be sufficient. Secondly, the data sampling X should also contain sufficient noiseless data to achieve robust capability. In fact, the first one can be easily met but the second one not. Because gene expression data are usually noisy, in reality, function (9) may be invalid and not robust.

To solve the problem in (9), we introduce the following LRR problem [20]:

$$\min_Z \|Z\|_*, \quad \text{s.t. } X_O = [X_O, X_H] Z, \tag{10}$$

where X_O is the observed data matrix and the X_H is the unobserved data, that is, the hidden data. We use the concatenation of X_O and X_H as a dictionary. The optimal result of (10) is $Z_{O,H}^* = [Z_{O|H}^*; Z_{H|O}^*]$, where $Z_{O|H}^*$ and $Z_{H|O}^*$ correspond to X_O and X_H, respectively.

By solving (10), the two problems above can be solved well. Then our next mission is to recover the affinity matrix $Z_{O|H}^*$ by using only X_O in the absence of the hidden data X_H. The method is called latent low-rank representation (LatLRR), which is an improvement of LRR.

Supposing we have two matrices X_O and X_H, then by solving (10) we have the following equations:

$$Z_{O|H}^* = V_O V_O^T, \qquad Z_{H|O}^* = V_H V_O^T, \tag{11}$$

where V_H and V_O can be obtained through computing the skinny singular value decomposition of $[X_O, X_H] = U \sum V^T$, and $V = [V_O; V_H]$. Namely, $X_O = U \sum V_O^T$ and $X_H = U \sum V_H^T$.

Depending on function (11), we have

$$\begin{aligned}
X_O &= [X_O, X_H] Z_{O,H}^* \\
&= X_O Z_{O|H}^* + X_H Z_{H|O}^* \\
&= X_O Z_{O|H}^* + X_H V_H V_O^T \\
&= X_O Z_{O|H}^* + U \sum V_H^T V_H V_O^T \\
&= X_O Z_{O|H}^* + U \sum V_H^T V_H \sum^{-1} U^T X_O.
\end{aligned} \tag{12}$$

Let $L_{H|O}^* = U \sum V_H^T V_H \sum^{-1} U^T$; then we have the following simple function:

$$X_O = X_O Z_{O|H}^* + L_{H|O}^* X_O. \tag{13}$$

If X_O and X_H come from the same collection of low-rank subspaces, then both $Z_{O|H}^*$ and $L_{H|O}^*$ should be of low-rank, so we can achieve

$$\min_{Z_{O|H}, L_{H|O}} \text{rank} \left(Z_{O|H} \right) + \text{rank} \left(L_{H|O} \right)$$

$$\text{s.t.} \quad X_O = X_O Z_{O|H} + L_{H|O} X_O. \tag{14}$$

Just as in [28–30], we also change the above rank minimization problem to the nuclear norm. Then we have the following convex optimization problem:

$$\min_{Z,L} \|Z\|_* + \|L\|_* \quad \text{s.t. } X = XZ + LX. \tag{15}$$

Here, we replace X_O, $Z_{O|H}$, and $L_{H|O}$ with X, Z, and L, respectively, for ease of representation. In (15), X is the noiseless observed data. By considering there may exist corrupted data or noise in X, we also need to introduce a denoising model about (15); then we have

$$\min_{Z,L} \|Z\|_* + \|L\|_* + \lambda \|E\|_1 \quad \text{s.t. } X = XZ + LX + E, \tag{16}$$

where $\lambda > 0$ is a scalar and $\|E\|_1$ is the l_1-norm of sparse noise matrix E. If $\lambda \rightarrow +\infty$, the problem (16) will be equivalent to (15), that is, no noise in the observed data X. In (16), the optimal solutions XZ^*, $L^* X$, and E^* represent the principal features, salient features, and noise, respectively.

To solve the LatLRR problem listed in (16), we introduce the augmented Lagrange multiplier (ALM) [33] method and revise (16) as follows to meet the requirement of ALM algorithm:

$$\min_{Z,L,J,S,E} \|Z\|_* + \|L\|_* + \lambda \|E\|_1 \quad \text{s.t. } X = XZ + LX + E,$$
$$Z = J, \quad L = S. \tag{17}$$

This problem can be solved by ALM method which minimizes the following augmented Lagrange function:

$$\begin{aligned}
&\|J\|_* + \|S\|_* + \lambda \|E\|_1 + \text{tr} \left(Y_1^T \left(X - XZ - LX - E \right) \right) \\
&\quad + \text{tr} \left(Y_2^T \left(Z - J \right) \right) + \text{tr} \left(Y_3^T \left(L - S \right) \right) \\
&\quad + \frac{\mu}{2} \left(\|X - XZ - LX - E\|_F^2 + \|Z - J\|_F^2 + \|L - S\|_F^2 \right),
\end{aligned} \tag{18}$$

where $\text{tr}(\cdot)$ and $\|\cdot\|_F$ denote the trace and Frobenius norm of a matrix, respectively. $\mu > 0$ is a penalty parameter. More details about (18) can be found in [33].

2.3. Sparse Representation Classification Based on LatLRR.
Since LatLRR can extract the salient features and remove noise from original data sets, in this study, before using observed data for classification, we firstly use LatLRR to suppress noise and get the salient features. Then we use the denoised data for tumor classification; that is, we factorize the observed data X into

$$X = XZ + LX + E. \tag{19}$$

Here, we only use $D = LX$ for data classification. For a test sample y, we can calculate its SR by the following function:

$$J(x, \lambda) = \min_x \left\{ \|Dx - Ly\|_2 + \lambda \|x\|_1 \right\}, \tag{20}$$

where the parameter $\lambda > 0$ can be determined experimentally and x is a coefficient vector. Assuming the test sample y

belongs to one of target classes, the training data set is sufficient. When classifying y, we introduce Ly, where L is a square matrix obtained through LatLRR method when extracting the salient features.

Ideally, Ly can be linearly represented by the samples from the same class in D. Namely, the representation vector x should be sparse and the nonzero entries are associated with the columns of D from the same class. This will lead us to classify the test samples. However, noise and modeling errors will also introduce some nonzero entries to x which correspond to the columns of D from the multiple classes [17]. To solve this problem, we classify Ly based on how well it can be reconstructed by using the coefficients from each class as in [17].

Using the result of (20), we construct $\delta_i(x)$ as the characteristic function which selects the coefficients associated with the ith class in the coefficient vector x. By only using ith class coefficients to reconstruct the test sample Ly as $\widehat{y}_i = D\delta_i(x)$, we can classify Ly into the minimum residual class between Ly and $\widehat{y}$; that is,

$$\min_i r_i(y) = \|Ly - D\delta_i(x)\|_2. \tag{21}$$

Our classification algorithm can be summarized as follows.

> *Input.* Observed data $X \in R^{m \times n}$ for k classes; test sample y.
>
> > *Step 1.* Normalize the columns of X.
> > *Step 2.* Extract the salient features of X and remove to some extent noise to get data D defined in (19).
> > *Step 3.* Solve the optimization problem defined in (20).
> > *Step 4.* Compute the residuals $r_i(y) = \|Ly - D\delta_i(x)\|_2$.
>
> *Output.* Identity$(y) = \arg \min_i r_i(y)$.

Our method can be seen as the combination of SRC [17] and latent low-rank representation for feature extraction [20], so we named it as SRC-LatLRR. In SRC, the test sample is represented as a sparse linear combination of the training samples from the same class. In LatLRR, noise is removed to some extent and salient features are simultaneously extracted from the training samples. So the introduction of LatLRR can improve the classification accuracy of SRC in a way.

2.4. Evaluation of the Performance. To evaluate our proposed method, we compare our method with SRC [17, 34], LASSO [35], and SVM [8, 36, 37]. SVM has been proved to be one of the best classifiers for classifying data in the area of "high dimensionality and small sample size" [36, 37]. We do binary classification and multiclass classification experiments in Sections 3.1 and 3.2, respectively. During the experiment, the best results of SRC, LASSO, and SVM are also used to compare with those of our method, which were achieved by choosing appropriate parameters experimentally. As the number of tumor sample is too small, we use stratified 10-fold cross validation in all our experiments. In the multiclass

TABLE 1: Three binary data sets used in the experiments.

Datasets	Samples		Genes
	Class 1	Class 2	
Colon cancer	40	22	2000
Prostate cancer	77	59	12600
DLBCL	58	19	5469

TABLE 2: Classification accuracies by different methods for the three binary data sets.

Datasets	SVM	LASSO	SRC	SRC-LatLRR
Colon cancer	85.48	85.48	85.48	**90.32**
Prostate cancer	91.18	91.91	**94.85**	94.12
DLBCL	96.10	96.10	**97.40**	**97.40**

classification experiments, we do not use LASSO method because it is designed only for binary class classification problems [35]. As we know, dimensionality reduction can improve the classification performance and computing speed, so we reduce data dimensionality using between-category to within-category sums of squares methods in our experiments.

3. Experimental Results

3.1. Two-Class Classification Problem. In this subsection, three two-class microarray data sets are used to evaluate our method: colon cancer [38], prostate cancer [39], and diffuse large B-cell lymphoma [40].

The colon data set contains 62 samples consisting of 40 tumor and 22 normal. The prostate data set contains prostate tumors and normal prostate samples, each consisting of the expression levels of 12600 genes. For the DLBCL data set, the gene expression values were measured by high-density oligonucleotide microarrays. An overview of the three data sets is given in Table 1.

The classification results by using SVM, LASSO, SRC, and the proposed SRC-LatLRR are listed in Table 2. From Table 2, we can see that our method SRC-LatLRR performs well on all the three data sets. Even the performance of SRC-LatLRR is not better than SRC on the prostate cancer data set, but it is better than SVM and LASSO. In summary, SRC has an advantage for the prostate cancer and DLBCL data sets, but SRC-LatLRR is the best classifier for the colon cancer and DLBCL data sets.

To further evaluate our method, in this experiment, we also introduced BW feature selection in our method to classify these three data sets. The results are listed in Table 3, and the number of genes selected is given in the parenthesis behind data set. From Table 3, we can see that after feature selection, our proposed classification method outperforms the other three classification methods, and it can even achieve an accuracy of 100% for the DLBCL data set.

3.2. Multiclass Classification Problem. In this subsection, we use four multiclass data sets to further check the classification

Table 3: Classification accuracies by different methods with gene selection for the three binary data sets.

Datasets	SVM	LASSO	SRC	SRC-LatLRR
Colon cancer (1000)	87.1	87.1	87.1	**91.94**
Prostate cancer (1500)	94.85	91.18	95.59	**96.32**
DLBCL (800)	97.40	93.51	97.40	**100**

Table 4: Descriptions of the four multiclass data sets used in DNA classification experiments.

Dataset	Class counts	Samples	Genes
Lung cancer	5	203	12600
Leukemia	3	72	11225
11_tumors	11	174	12533
9_tumors	9	60	5726

Table 5: Classification accuracies by different methods for the multiclass data sets.

Dataset	SVM	SRC	SRC-LatLRR
Lung cancer	**96.05**	95.07	95.07
Leukemia	96.60	95.83	**98.61**
11_tumors	94.68	**94.83**	**94.83**
9_tumors	65.10	**66.67**	**66.67**

performance of SRC-LatLRR. The four data sets are lung cancer [41], leukemia [42], 11_tumors [43], and 9_tumors [44].

In lung cancer data set, there are four classes of lung cancer and normal class. This data set contains 203 samples. For leukemia data set, all the samples are classified into acute myelogenous leukemia, acute lymphoblastic leukemia, or mixed-lineage leukemia. The data set includes 72 samples with 11225 genes. For 11_tumors, there are 11 classes of samples, which are ovary, bladder/ureter, breast, colorectal, gastroesophagus, kidney, liver, prostate, pancreas, adeno lung, and squamous lung. This data set includes 174 samples. For the 9_tumors data set, there are 60 samples with 5726 genes. These 9 types of tumors are non-small-cell lung, colon, breast, ovarian, leukemia, renal, melanoma, prostate, and central nervous system. The detailed descriptions about these four data sets are listed in Table 4. All the four data sets were produced by oligonucleotide microarrays and the analysis tool Affymetrix GENECHIP [36].

The experimental results are listed in Table 5. From these results, we can see that the proposed method SRC-LatLRR does not have a clear advantage over SVM and SRC. The reason may be that in these data sets, the training samples of each class are very few so that the sample space is not complete.

We then introduced BW feature selection before applying our method. The obtained results are listed in Table 6. From the results we can see that the proposed method classified leukemia well. For the other data sets, it has no clear

Table 6: Classification accuracies by different methods with gene selection for the multiclass data sets.

Dataset	SVM	SRC	SRC-LatLRR
Lung cancer (2000)	**96.62**	95.07	95.57
Leukemia (3000)	96.90	95.83	**98.61**
11_tumors (1000)	**96.07**	95.40	95.40
9_tumors (2000)	**85.84**	71.67	80.00
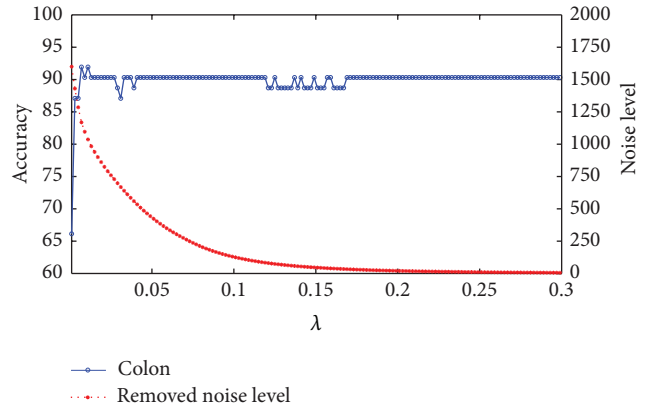
Figure 1: The changing curves of classification accuracy and removed noise level with λ on the colon data set.

advantage. But it performed better than SRC for all the four data sets.

3.3. The Choice of the Balanced Parameter. In this section, we use the data sets described in Section 3.1 to check how λ in (16) affect the classification performance. We show the accuracies and the removed noise level by our method at different values of λ in Figures 1, 2, and 3 for the colon, prostate, and DLBCL data sets, respectively. From (16), we know that the lower the λ is, the bigger the noise level is removed. For these three figures we use $\|E\|_1$ to represent the level of the removed noise. From these three figures we can see that the noise that we remove from the original data can not be too much, or it will reduce the accuracy. The reason is that if λ is set to be too small, useful information may be also removed besides noise. On the contrary, if λ is too big, the noise that was removed is too little, and we still can not get a good classification result. The experiment suggests that for colon data sets, $\lambda = 0.011$ is the best choice and $\lambda = 0.096$ and $\lambda = 0.1$ for the prostate and DLBCL data sets, respectively.

4. Conclusions

For gene expression data, cancer diagnosis is one of the most important clinical applications. In this paper, we have proposed a new SR-based method for tumor classification which uses the noiseless salient features extracted from the original samples to classify a test sample. We compared our method with several state-of-the-art methods including SVM, LASSO, and SRC on seven data sets. The results of experiments show that the proposed method is better than
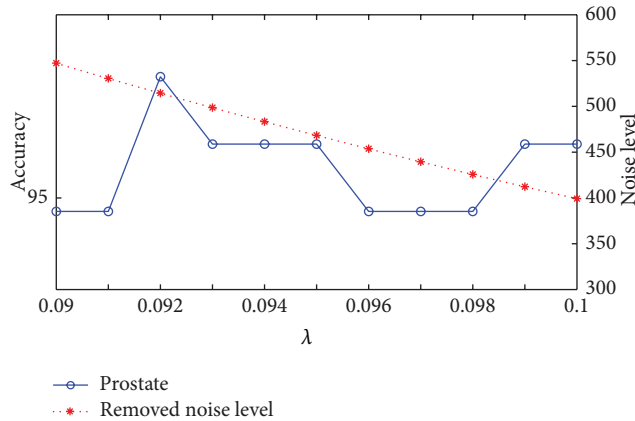
FIGURE 2: The changing curves of classification accuracy and removed noise level with λ on the prostate data set.
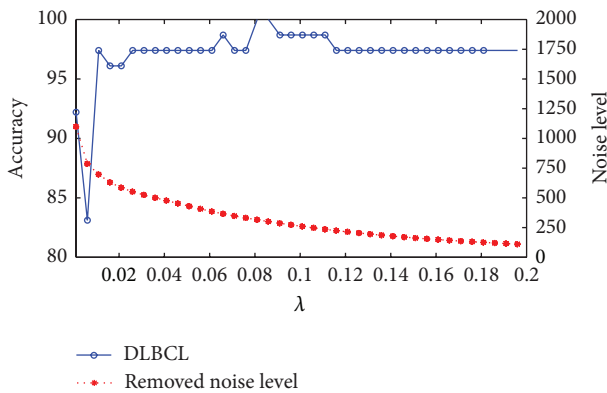

FIGURE 3: The changing curves of classification accuracy and removed noise level with λ on the DLBCL data set.

SVM, LASSO, and SRC in a way. These demonstrate that SRC-LatLRR is effective and efficient for tumor classification. We also introduced gene selection into our method. The results show that gene selection can improve the classification accuracy to some extent.

During the study we also found that, for the optimal result of LatLRR on the observed samples, Z^* represents the affinity matrix of samples [21]. In theory, the affinity matrix can be used to cluster samples. In future, we will extend it to investigate the property of sample clusters.

Abbreviations

SR: Sparse representation
SRC: Sparse representation classification
LRR: Low-rank representation
LatLRR: Latent low-rank representation
ALM: Augmented Lagrange multiplier
SVM: Support vector machines
LASSO: Least absolute shrinkage and selection operator
BW: Between-categories to within-category.

References

[1] A. Cánovas, G. Rincon, A. Islas-Trejo, S. Wickramasinghe, and J. F. Medrano, "SNP discovery in the bovine milk transcriptome using RNA-Seq technology," *Mammalian Genome*, vol. 21, no. 11-12, pp. 592–598, 2010.

[2] T. R. Golub, D. K. Slonim, P. Tamayo et al., "Molecular classification of cancer: class discovery and class prediction by gene expression monitoring," *Science*, vol. 286, no. 5439, pp. 531–527, 1999.

[3] J. P. Brunet, P. Tamayo, T. R. Golub, and J. P. Mesirov, "Metagenes and molecular pattern discovery using matrix factorization," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 12, pp. 4164–4169, 2004.

[4] Y. Gao and G. Church, "Improving molecular cancer class discovery through sparse non-negative matrix factorization," *Bioinformatics*, vol. 21, no. 21, pp. 3970–3975, 2005.

[5] D. S. Huang and C. H. Zheng, "Independent component analysis-based penalized discriminant method for tumor classification using gene expression data," *Bioinformatics*, vol. 22, no. 15, pp. 1855–1862, 2006.

[6] T. K. Paul and H. Iba, "Prediction of cancer Class with majority voting genetic programming classifier using gene expression data," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 6, no. 2, pp. 353–367, 2009.

[7] C. H. Zheng, L. Zhang, T. Y. Ng, C. K. Shiu, and D. S. Huang, "Metasample-based sparse representation for tumor classification," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 8, no. 5, pp. 1273–1282, 2011.

[8] T. S. Furey, N. Cristianini, N. Duffy, D. W. Bednarski, M. Schummer, and D. Haussler, "Support vector machine classification and validation of cancer tissue samples using microarray expression data," *Bioinformatics*, vol. 16, no. 10, pp. 906–914, 2000.

[9] Y. S. Lee, A. Krishnan, Q. Zhu, and O. G. Troyanskaya, "Ontology-aware classification of tissue and cell-type signals in gene expression profiles across platforms and technologies," *Bioinformatics*, vol. 29, no. 23, pp. 3036–3044, 2013.

[10] M. Tanic, E. Andress, S. M. Rodriguez-Pinilla et al., "MicroRNA-based molecular classification of non-BRCA1/2 hereditary breast tumours," *British Journal of Cancer*, vol. 109, no. 10, pp. 2724–2734, 2013.

[11] J. H. Huang, H. L. Xie, J. Yan, H. M. Lu, Q. S. Xu, and Y. Z. Liang, "Rsing random forest to classify T-cell epitopes based on amino acid properties and molecular features," *Analytica Chimica Acta*, vol. 804, pp. 70–75, 2013.

[12] L. Nanni, S. Brahnam, S. Ghidoni, E. Menegatti, and T. Barrier, "Acomparison of methods for extracting information from

the co-occurrence matrix for subcellular classification," *Expert Systems with Applications*, vol. 40, no. 18, pp. 7457–7467, 2013.

[13] G. R. Lioyd, L. M. Almond, N. Stone et al., "Utilising non-consensus pathology measurements to improve the diagnosis of oesophageal cancer using a raman spectroscopic probe," *The Analyst*, vol. 139, no. 2, pp. 381–388, 2014.

[14] G. Braz, S. V. da Rocha, M. Gattass, A. C. Silva, and A. C. de Paiva, "A mass classification using spatial diversity approaches in mammography images for false positive reduction," *Expert Systems with Applications*, vol. 40, no. 18, pp. 7534–7543, 2013.

[15] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Review*, vol. 43, no. 1, pp. 129–159, 2001.

[16] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[17] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 2, pp. 210–227, 2009.

[18] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman, "Supervised dictionary learning," in *Proceedings of the 22nd Annual Conference on Neural Information Processing Systems (NIPS '09)*, 2009.

[19] G. C. Liu, Z. C. Lin, S. C. Yan, J. Sun, Y. Yu, and Y. Ma, "Robust recovery of subspace structures by low-rank representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 1, pp. 171–184, 2013.

[20] G. Liu and S. C. Yan, "Latent low-rank representation for subspace segmentation and feature extraction," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV '11)*, pp. 1615–1622, Barcelona, Spain, November 2011.

[21] G. C. Liu, Z. C. Lin, and Y. Yu, "Robust subspace segmentation by low-rank representation," in *Proceedings of the 27th International Conference on Machine Learning (ICML '10)*, pp. 663–670, 2010.

[22] G. C. Liu, H. Xu, and S. C. Yan, "Exact subspace segmentation and outlier detection by low-rank representation," *JMLR: Workshop and Conference Proceedings*, vol. 22, pp. 703–711, 2012.

[23] L. Ma, C. H. Wang, B. H. Xiao, and W. Zhou, "Sparse representation for face recognition based on discriminative low-rank dictionary learning," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '12)*, pp. 2586–2593, June 2012.

[24] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 489–509, 2006.

[25] E. J. Candes and T. Tao, "Near-optimal signal recovery from random projections: universal encoding strategies?" *IEEE Transactions on Information Theory*, vol. 52, no. 12, pp. 5406–5425, 2006.

[26] R. Tibshirani, "Regression shrinkage and selection via the lasso: a retrospective," *Journal of the Royal Statistical Society B*, vol. 73, no. 3, pp. 273–282, 2011.

[27] S. J. Kim, K. Koh, M. Lustig, S. Boyd, and D. Gorinevsky, "An interior-point method for large-scale l_1-regularized least squares," *IEEE Journal on Selected Topics in Signal Processing*, vol. 1, no. 4, pp. 606–617, 2007.

[28] E. J. Candes and Y. Plan, "Matrix completion with noise," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 925–936, 2010.

[29] E. J. Candès and B. Recht, "Exact matrix completion via convex optimization," *Foundations of Computational Mathematics*, vol. 9, no. 6, pp. 717–772, 2009.

[30] R. H. Keshavan, A. Montanari, and S. Oh, "Matrix completion from noisy entries," *Journal of Machine Learning Research*, vol. 11, pp. 2057–2078, 2010.

[31] M. Fazel, *Matrix rank minimization with applications [Ph.D. thesis]*, Stanford University, Stanford, Calif, USA, 2002.

[32] Y. Ni, J. Sun, X. Yuan, S. Yan, and L. F. Cheong, "Robust low-rank subspace segmentation with semidefinite guarantees," in *Proceedings of the 10th IEEE International Conference on Data Mining Workshops (ICDMW '10)*, pp. 1179–1188, Sydney, Australia, December 2010.

[33] Z. Lin, M. Chen, L. Wu, and Y. Ma, "The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices," Tech. Rep. UILU-ENG-09-2215, 2009.

[34] X. Hang and F. X. Wu, "Sparse representation for classification of tumors using gene expression data," *Journal of Biomedicine and Biotechnology*, vol. 2009, Article ID 403689, 6 pages, 2009.

[35] D. Ghosh and A. M. Chinnaiyan, "Classification and selection of biomarkers in genomic data using LASSO," *Journal of Biomedicine and Biotechnology*, vol. 2005, no. 2, pp. 147–154, 2005.

[36] A. Statnikov, C. F. Aliferis, I. Tsamardinos, D. Hardin, and S. Levy, "A comprehensive evaluation of multicategory classification methods for microarray gene expression cancer diagnosis," *Bioinformatics*, vol. 21, no. 5, pp. 631–643, 2005.

[37] N. Pochet, F. de Smet, J. A. K. Suykens, and B. L. R. de Moor, "Systematic benchmarking of microarray data classification: assessing the role of non-linearity and dimensionality reduction," *Bioinformatics*, vol. 20, no. 17, pp. 3185–3195, 2004.

[38] U. Alon, N. Barka, D. A. Notterman et al., "Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 96, no. 12, pp. 6745–6750, 1999.

[39] D. Singh, P. G. Febbo, K. Ross et al., "Gene expression correlates of clinical prostate cancer behavior," *Cancer Cell*, vol. 1, no. 2, pp. 203–209, 2002.

[40] M. A. Shipp, K. N. Ross, P. Tamayo et al., "Diffuse large B-cell lymphoma outcome prediction by gene-expression profiling and supervised machine learning," *Nature Medicine*, vol. 8, no. 1, pp. 68–74, 2002.

[41] A. Bhattacharjee, W. G. Richards, J. Staunton et al., "Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 98, no. 24, pp. 13790–13795, 2001.

[42] S. A. Armstrong, J. E. Staunton, L. B. Silverman et al., "MLL translocations specify a distinct gene expression profile that distinguishes a unique leukemia," *Nature Genetics*, vol. 30, no. 1, pp. 41–47, 2002.

[43] A. I. Su, J. B. Welsh, L. M. Sapinoso et al., "Molecular classification of human carcinomas by use of gene expression signatures," *Cancer Research*, vol. 61, no. 20, pp. 7388–7393, 2001.

[44] J. E. Staunton, D. K. Slonim, H. A. Coller et al., "Chemosensitivity prediction by transcriptional profiling," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 98, no. 19, pp. 10787–10792, 2001.