A Parallel Gaussian–Bernoulli Restricted Boltzmann Machine for Mining Area Classification With Hyperspectral Imagery

Kun Tan, Senior Member, IEEE, Fuyu Wu, Qian Du, Fellow, IEEE, Peijun Du, Senior Member, IEEE, and Yu Chen

Abstract—In this paper, a novel feature extraction method is proposed for hyperspectral image classification using a Gaussian–Bernoulli restricted Boltzmann machine (GBRBM) in parallel. The proposed approach employs several GBRBMs with different hidden layers to extract deep features from hyperspectral images, which are nonlinear and local invariant. Based on the learned deep features, a logistic regression layer is trained for classification. The proposed approaches are carried out on two public hyperspectral datasets: Pavia University dataset and Salinas dataset, and a new dataset obtained by HySpex imaging spectrometer in the mining area in Xuzhou. The obtained results reveal that the proposed approach offers superior performance compared to traditional classifiers. The advantage of the proposed GBRBM is that it can extract deep features in an unsupervised way and reduce the prediction time by using GPU. In particular, the classification results of the mining area provide valuable suggestions to improve environmental protection.

Index Terms—Deep learning, Gaussian–Bernoulli restricted Boltzmann machine (GBRBM), hyperspectral image classification.

I. INTRODUCTION

HYPERSPECTRAL image classification has important remote sense applications. Many machine learning techniques have been applied, such as K-nearest neighbor [1], Bayes classifier [2], and support vector machine (SVM) [3], to achieve classification. However, the increase of spectral dimension leads to low efficiency of classification algorithms and large computational cost [4]. Therefore, it is important to reduce dimensionality before classification. Feature extraction is one of dimensionality reduction methods by transforming the original features into a lower dimensional feature space. The commonly used linear models include the independent component analysis [5], linear discriminant analysis [6], and their variants. In [7], Wen et al. presented a supervised linear manifold learning feature extraction method for hyperspectral image classification. In [8], Friedman et al. alternated the usual maximum likelihood estimates for the covariance matrices and proposed the regularized discriminant analysis, and quadratic discriminant analysis. In [9], Howland et al. modified the generalized singular value decomposition for the commonly used SVD algorithm to solve the small-sample-size problem. However, the linear models cannot deal with high-dimensional data effectively. In order to solve these problems, a variety of nonlinear models, such as locally linear embedding [10], Isomap [11], and Laplacian eigenmaps [12], were proposed. The nonlinear models can effectively exploit the intrinsic structure of data. In [13], Wu et al. introduced the kernel function into the PCA algorithm, which was applied to the near-infrared spectral data for feature extraction. In [14], Sun et al. proposed a fast and robust principal component analysis on the Laplacian graph for hyperspectral band selection, which regularized the Laplacian graph into a regular robust principal component analysis. In [15], Mika et al. presented a Fisher discriminant analysis combined with a kernel function, but the disadvantage is that the model is complex and time-consuming. Sun et al. proposed a sparse and low-rank near-isometric linear embedding to extract features in hyperspectral imagery, which performed the best among comparative methods [16].

Recently, deep learning has made a major breakthrough in all aspects [17]. For instance, a deep neural network (DNN) can extract high-level features for efficient classification [18]. Currently, the DNN mainly includes deep belief networks (DBN) [19], deep Boltzmann machines [20], deep auto-encoder neural networks [21], and deep convolutional neural networks [22]. They also provide a new solution for feature extraction and classification of hyperspectral images. In [23], Zhang et al. gave a technical tutorial of deep learning for the remote sensing data analysis, which introduced the applications of deep learning in
different perspectives of image preprocessing, such as pixel-based classification, target recognition, and scene understanding. In [24], a DBN-based classifier achieved a spectral–spatial classification of hyperspectral data. In [25], Lin et al. applied the deep auto-encoder neural networks for the classification of hyperspectral imagery. Meanwhile, many methods combining traditional classification strategies with deep learning were proposed. In [26], Ma et al. presented a novel semisupervised classification based on multi-decision labeling and deep feature learning. In [27], Wang et al. introduced a hybrid of PCA, guided filtering, and deep learning into hyperspectral image classification. Recently, multiple neural networks have been shown to be successful in computer vision. In [28], Farabet et al. introduced a discriminative framework using a multi-scale convolutional network operated on raw pixels to learn appropriate low-level and mid-level features for scene parsing. In [29], Längkvist et al. found that multiple CNNs in parallel with varying contest can achieve stable classification accuracy. In [30], Zhou et al. raised a group belief network based on unlabeled hyperspectral data, which was the first attempt to incorporate group knowledge of hyperspectral features for classification.

However, in the application of hyperspectral classification, it has been noticed that classification accuracy may be degraded with the increase of hidden layers. There are two main reasons for this phenomenon: first, the depth of DNN directly affects the number of parameters, and the deeper the DNN, the more the number of parameters. Thus, due to lack of training samples, the DNN cannot be trained effectively. Second, the classification of hyperspectral images is carried out pixel by pixel, which is different from natural images. The number of bands in hyperspectral images is about 200, which leads to the absence of DNNs for feature extraction. It has been argued that a single-hidden-layer network with well-tuned parameters may be more effective than DNN [31].

In this paper, we focus on single-layer Gaussian–Bernoulli restricted Boltzmann machines (GBRBM) and multiple GBRBMs in parallel to extract the features of hyperspectral data. The extracted features are then used by a logistic regression (LR) layer to address the classification problem. Compared with the contrast methods, the parallel GBRBM method can make full use of spectral information by combining different features. Moreover, this method has fast convergence due to fewer parameters. According to experiments, the parallel GBRBM can offer better classification accuracy with lower computational cost, which is more suitable for applications of hyperspectral datasets.

The rest of the paper is organized as follows. The theory of GBRBM and the structure of GBRBM in parallel are given in Section II. Datasets and experimental results are shown in Section III. An application of the model to a mining area is given in Section IV, and conclusions are drawn in Section V.

II. PROPOSED METHOD

A. Restricted Boltzmann Machine (RBM)

An RBM is a two-layer network, which contains a “visible” layer \( v = [0, 1]^D \) and a “hidden” layer \( h = [0, 1]^F \). The illustration of RBM is shown in Fig. 1.

A joint configuration \((v, h)\) of the visible and hidden layers has an energy given by [32]

\[
E(v, h; \theta) = - \sum_{i=1}^{D} b_i v_i - \sum_{j=1}^{F} a_j h_j - \sum_{i=1}^{D} \sum_{j=1}^{F} w_{ij} v_i h_j
\]

where \( \theta = [a_j, b_i, w_{ij}] \), \( w_{ij} \) represents the weight between the hidden and visible layers, and \( a_j \) and \( b_i \) are bias terms of the hidden and visible layers, respectively. The joint distribution over the layers is defined by

\[
P(v, h; \theta) = \frac{1}{Z(\theta)} \exp(-E(v, h; \theta))
\]

where the “partition function” \( Z(\theta) \) is given by summing over all possible pairs of hidden and visible vectors as

\[
Z(\theta) = \sum \sum E(v, h; \theta).
\]

The conditional distributions of the hidden layer \( h \) and visible layer \( v \) are given by

\[
p(h_j = 1|v) = \sigma(a_j + \sum_i v_i W_{ij})
\]

\[
p(v_i = 1|h) = \sigma(b_i + \sum_j h_j W_{ij})
\]

\[
\sigma(x) = \frac{1}{1 + \exp(x)}.
\]

In general, a method named contrastive divergence (CD) is used to train the RBM model. The model of CD is illustrated in Fig. 2.
During training, the parameters of RBM can be updated as

\[ w_{ij} = w_{ij} + \varepsilon_a \left( E_{\text{data}} [v_i h_j] - E_{\text{model}} [v_i h_j] \right) \]

\[ a_j = a_j + \varepsilon_a \left( E_{\text{data}} [h_j] - E_{\text{model}} [h_j] \right) \]

\[ b_i = b_i + \varepsilon_b \left( E_{\text{data}} [v_i] - E_{\text{model}} [v_i] \right) \]

where \( \varepsilon \) is the learning rate, \( E_{\text{data}} \) is the expectation of visible layer in an active state, and \( E_{\text{model}} \) is the expectation of visible layer in any active states.

**B. Gaussian–Bernoulli Restricted Boltzmann Machines**

The conventional RBM requires the state of layers to follow binary distribution, which limits their application. To solve the problem, a popular approach is to replace the binary visible layer with Gaussian ones [33]. The modified model is named GBRBM whose structure is same as the RBM [34]. The energy function of GBRBM is given by

\[ E(v, h) = \sum_{i=1}^{V} \left( \frac{(v_i - b_i)^2}{2\sigma^2} - \sum_{j=1}^{H} c_j h_j - \sum_{i=1}^{V} \sum_{j=1}^{H} \frac{v_i}{\sigma^2} h_j w_{ij} \right). \]

(10)

Unlike the traditional RBM, the conditional distributions of the hidden layer \( h \) and visible layer \( v \) are

\[ p(v|h) = \frac{e^{-E(v,h)}}{\int e^{-E(v,h)} du} \sim N \left( b_i + \sigma_i \sum_{j=1}^{H} h_j w_{ij}, \sigma^2 \right) \]

\[ p(h_j = 1|v) = \frac{1}{1 + e^{(\sum_{i=1}^{V} \frac{v_i}{\sigma^2} w_{ij} + c_j)}} = \sigma \left( \sum_{i=1}^{V} \frac{v_i}{\sigma^2} w_{ij} + c_j \right). \]

(11)

(12)

During model training, the parameters of GBRBM are updated as

\[ w_{ij} = w_{ij} + \varepsilon_w \left( E_{\text{data}} \left[ \frac{v_i h_j}{\sigma^2} \right] - E_{\text{model}} \left[ \frac{v_i h_j}{\sigma^2} \right] \right) \]

\[ c_j = c_j + \varepsilon_b \left( E_{\text{data}} [h_j] - E_{\text{model}} [h_j] \right) \]

\[ b_i = b_i + \varepsilon_b \left( E_{\text{data}} \left[ \frac{v_i}{\sigma^2} \right] - E_{\text{model}} \left[ \frac{v_i}{\sigma^2} \right] \right) \]

\[ \sigma_i = \sigma_i + \varepsilon_{\sigma} \left( E_{\text{data}} \left( \frac{(v_i - b_i)^2}{\sigma^2} - \sum_{j=1}^{H} h_j w_{ij} \right) \right) - E_{\text{model}} \left( \frac{(v_i - b_i)^2}{\sigma^2} - \sum_{j=1}^{H} h_j w_{ij} \right). \]

(13)

(14)

(15)

(16)

The parameter \( \sigma \) has an influence on the gradient calculation of other parameters in the process of model optimization. In general, to avoid this problem, the standardization pretreatment of the input data should be conducted before feature extraction. In this research, the parameter \( \sigma \) is set to be a constant 1, and it does not participate in the optimization process to reduce model complexity.

After pre-processing, the GBRBM can extract features from spectral information in an unsupervised way. In this paper, the features are used as the input data for classification using LR. The entire structure is shown in Fig. 3.

**C. Per-Pixel Classification Using Multiple GBRBM in Parallel**

Recently, multiple feature learning has acquired better results in the field of computer vision. The features extracted by a single GBRBM model are always limited, and the combination of multiple features extracted by several GBRBM models is more powerful for classification. Thus, in this paper, a multiple feature learning based on GBRBM is proposed. The basic idea is to run several GBRBM models with varying hidden neurons in parallel, and then to combine the output from each model to the LR classifier. The whole structure is shown in Fig. 4. Each GBRBM model is trained and features are extracted separately in an unsupervised way. The features extracted from different models are combined together for the LR classification. Finally, the class labels are calculated in a supervised way by the LR. Another possibility of achieving multiscale feature learning is to use the same hidden neurons for GBRBM model, but increasing the depth of the GBRBM models. However, according to previous studies, in the application of hyperspectral classification, it has been noticed that classification accuracy may be degraded with the increase of hidden layers. Thus, we choose to increase the number of hidden neurons in this paper.

**III. EXPERIMENTS WITH STANDARD DATA**

In this section, we evaluate our method using standard hyperspectral data. Experiments were implemented based on the deep learning framework Theano and executed on a PC with an Intel single Core i5 CPU, NVIDIA GTX-1070 GPU.
A. Dataset Description

In this paper, two standard hyperspectral datasets (i.e., Pavia University and Salinas dataset) are used to verify the effectiveness of the proposed model. The Pavia University dataset, as shown in Fig. 5, acquired by the ROSIS sensor during a flight campaign over Pavia. This dataset has $610 \times 340$ pixels and 103 spectral bands. There are nine classes in this dataset. The Salinas dataset with 16 classes, as shown in Fig. 6, was collected by the 224-band AVIRIS sensor over Salinas Valley, California, and is characterized by high spatial resolution.

To evaluate the classification accuracy, 10% of labeled samples were selected as training samples and rest for testing. The numbers of training and testing samples for each class are given in Tables I and II. As SVM and DBN have been widely used in the application of remote sensing classification, they were introduced for comparison in this study. Otherwise, multi-grained cascade-forest (Gcforest) was a new DNN that has produced excellent performance in some datasets, so it was also introduced for comparison. In the experiment, each classifier was run 20 times with different training samples to avoid the fortuity, and then overall accuracy (OA), average accuracy (AA), kappa statistic, training, and prediction time were reported.

B. Spectral Classification Using GBRBM-LR Framework

In the GBRBM model, the number of hidden neurons is an important parameter, which affects the complexity of the model. Meanwhile, it also indicates the number of features to be extracted. In this section, the model with different numbers of hidden neurons are used. The results are shown in Tables III and IV. It can be seen that with the increase in hidden neurons, the OA and AA increased, and the kappa statistic also increased. However, the training and prediction time also increased.
neurons, classification accuracy is improved, but when the hidden neurons are increased to a certain level, classification accuracy becomes almost stable. According to these experiments, 100 hidden neurons can obtain satisfactory results.

Meanwhile, in order to verify the advantage of the GBRBM model, linear SVM, SVM with RBF kernel, DBN, and Gcforest are used for comparison. The number of hidden neurons is chosen to be 100 with the consideration of accuracy and efficiency. The accuracies over different classification methods are reported in Tables V and VI. Figs. 7 and 8 are classification maps of the model with different numbers of hidden neurons for the two datasets. The GBRBM model can extract the features of hyperspectral data effectively. However, the capability of a single GBRBM is still limited.

In this experiment, we use CUDA to make the GBRBM’s training procedure faster. Due to complexity, the proposed method takes longer training time compared to other algorithms, as shown in Tables V and VI. Although the parallel GBRBM-LR framework has a longer training time, its shorter prediction time makes it suitable for practical applications.

Fig. 7. Best classification results of the Pavia University dataset. (a) $H = 25$. (b) $H = 50$. (c) $H = 100$. (d) $H = 200$. (e) $H = 400$. (f) $H = 600$.

Fig. 8. Best classification results of the Salinas dataset. (a) $H = 25$. (b) $H = 50$. (c) $H = 100$. (d) $H = 200$. (e) $H = 400$. (f) $H = 600$.

### TABLE IV
**Classification With Different Number of Hidden Neurons From Salinas Dataset**

<table>
<thead>
<tr>
<th>Model</th>
<th>OA(%)</th>
<th>AA(%)</th>
<th>Kappa statistic</th>
<th>Training time(s)</th>
<th>Prediction time(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>H=25</td>
<td>93.50</td>
<td>93.54</td>
<td>0.9279</td>
<td>82.17</td>
<td>0.08</td>
</tr>
<tr>
<td>H=50</td>
<td>93.68</td>
<td>93.83</td>
<td>0.9297</td>
<td>82.24</td>
<td>0.10</td>
</tr>
<tr>
<td>H=100</td>
<td>93.81</td>
<td>93.87</td>
<td>0.9310</td>
<td>93.66</td>
<td>0.14</td>
</tr>
<tr>
<td>H=200</td>
<td>93.29</td>
<td>93.42</td>
<td>0.9253</td>
<td>93.84</td>
<td>0.15</td>
</tr>
<tr>
<td>H=400</td>
<td>93.83</td>
<td>93.88</td>
<td>0.9228</td>
<td>92.80</td>
<td>0.18</td>
</tr>
<tr>
<td>H=600</td>
<td>93.79</td>
<td>93.81</td>
<td>0.9317</td>
<td>97.49</td>
<td>0.23</td>
</tr>
</tbody>
</table>

### TABLE V
**Comparison Results of Classification From Pavia University Dataset (10%)**

<table>
<thead>
<tr>
<th>Model</th>
<th>OA(%)</th>
<th>AA(%)</th>
<th>Kappa statistic</th>
<th>Training time(s)</th>
<th>Prediction time(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GBRBM in Parallel</td>
<td>96.22</td>
<td>96.58</td>
<td>0.9491</td>
<td>238.93</td>
<td>0.29</td>
</tr>
<tr>
<td>SVM (Linear)</td>
<td>92.17</td>
<td>92.93</td>
<td>0.9128</td>
<td>0.41</td>
<td>5.34</td>
</tr>
<tr>
<td>SVM (RBF)</td>
<td>94.21</td>
<td>94.26</td>
<td>0.9420</td>
<td>0.53</td>
<td>10.33</td>
</tr>
<tr>
<td>DBN</td>
<td>92.39</td>
<td>91.38</td>
<td>0.8998</td>
<td>752.79</td>
<td>0.1292</td>
</tr>
<tr>
<td>Gcforest</td>
<td>90.67</td>
<td>90.76</td>
<td>0.8744</td>
<td>52.23</td>
<td>25.22</td>
</tr>
</tbody>
</table>

### TABLE VI
**Comparison Results of Classification From Salinas University Dataset (10%)**

<table>
<thead>
<tr>
<th>Model</th>
<th>OA(%)</th>
<th>AA(%)</th>
<th>Kappa statistic</th>
<th>Training time(s)</th>
<th>Prediction time(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GBRBM in Parallel</td>
<td>95.94</td>
<td>95.98</td>
<td>0.9543</td>
<td>355.69</td>
<td>0.38</td>
</tr>
<tr>
<td>SVM (Linear)</td>
<td>91.65</td>
<td>92.93</td>
<td>0.9180</td>
<td>0.67</td>
<td>15.48</td>
</tr>
<tr>
<td>SVM (RBF)</td>
<td>94.27</td>
<td>94.24</td>
<td>0.9361</td>
<td>1.14</td>
<td>32.71</td>
</tr>
<tr>
<td>DBN</td>
<td>91.74</td>
<td>95.30</td>
<td>0.9083</td>
<td>1400.33</td>
<td>0.12</td>
</tr>
<tr>
<td>Gcforest</td>
<td>92.41</td>
<td>95.70</td>
<td>0.9154</td>
<td>60.21</td>
<td>32.15</td>
</tr>
</tbody>
</table>
C. Spectral Classification Using GBRBM in Parallel-LR Framework

Based on the experimental results in the previous section, after the number of hidden neurons reaches 200, classification accuracy will not further advance but the calculation time is significantly increased. Thus, considering the calculation time and accuracy, the hidden neurons are selected as 25, 50, 100, or 200 in the parallel-LR framework. Figs. 9 and 10 show the classification accuracy for all possible combinations of using 1, 2, 3, or all 4 of the GBRBMs in parallel. It can be seen from Fig. 9 that the classification accuracy of Pavia University dataset using $25 + 50 + 100 + 200$ in parallel is 96.22%, which is higher than the accuracy using a single layer with 100 hidden neurons. Fig. 10 shows that the classification accuracy of Salinas dataset using $25 + 50$ in parallel is 95.44%, which is higher than the accuracy using a single layer with 100 hidden neurons. Therefore, it can be concluded that classification accuracy using GBRBM in parallel is higher than a single layer with more hidden neurons.

Meanwhile, for the two datasets, the network with the highest accuracy is also different. The Salinas dataset can achieve the highest accuracy (95.94%) by using three layers ($25 + 100 + 200$) in parallel. The Pavia University dataset achieves the highest accuracy (96.22%) by using four layers ($25 + 50 + 100 + 200$) in parallel. The accuracy using GBRBM in parallel is better than the accuracy from other models. The comparative result is shown in Tables V and VI. Figs. 11 and 12 illustrate the classification maps. The experiment shows that the GBRBM in parallel-LR framework can classify effectively hyperspectral data, and the classification accuracy will be improved with the increase of hidden neurons. However, when the hidden neurons are increased to a certain level, classification accuracy becomes stable.

Moreover, to verify the stability of the model using less training samples, 5% of labeled samples were selected as training and rest for testing. The comparative result is shown in Tables VII and VIII. Figs. 13 and 14 illustrate the classification maps. According to the results, reducing the number of training samples will degrade the accuracy of the GBRBM in the parallel model, DBN, and Gcforest. And reducing the number of training samples does not affect the SVM model. The reason for this phenomenon may be that the models have a large number of parameters to be trained, but SVM does not have such a problem.

IV. APPLICATION TO MINING AREA

While promoting the development of national economy, the coal industry has also brought about a series of environmental problems, such as soil quality decline, land subsidence, ecosystem degradation, etc. With the advance of spatial information
Fig. 12. Best classification results of the Salinas dataset (10%). (a) Labeled map. (b) GBRBM in parallel (25 + 100 + 200). (c) SVM (linear). (d) SVM (RBF). (e) DBN. (f) Gcforest.

Fig. 13. Best classification results of the Pavia University dataset (5%). (a) Labeled map. (b) GBRBM in parallel (25 + 50 + 100 + 200). (c) SVM (linear). (d) SVM (RBF). (e) DBN. (f) Gcforest.

Fig. 14. Best classification results of the Salinas dataset (5%). (a) Labeled map. (b) GBRBM in parallel (25 + 100 + 200). (c) SVM (Linear). (d) SVM (RBF). (e) DBN. (f) Gcforest.

TABLE VII

<table>
<thead>
<tr>
<th>Model</th>
<th>OA(%)</th>
<th>AA(%)</th>
<th>Kappa statistic</th>
<th>Training time(s)</th>
<th>Prediction time(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GBRBM in Parallel</td>
<td>92.20</td>
<td>90.93</td>
<td>0.8960</td>
<td>892.88</td>
<td>0.30</td>
</tr>
<tr>
<td>SVM (Linear)</td>
<td>86.82</td>
<td>79.53</td>
<td>0.8205</td>
<td>0.11</td>
<td>3.25</td>
</tr>
<tr>
<td>SVM (RBF)</td>
<td>93.13</td>
<td>92.67</td>
<td>0.9085</td>
<td>0.19</td>
<td>4.33</td>
</tr>
<tr>
<td>DBN</td>
<td>90.92</td>
<td>89.44</td>
<td>0.8741</td>
<td>407.15</td>
<td>0.10</td>
</tr>
<tr>
<td>Gcforest</td>
<td>89.12</td>
<td>88.96</td>
<td>0.8536</td>
<td>52.54</td>
<td>19.85</td>
</tr>
</tbody>
</table>

technology, remote sensing has been widely used in environmental monitoring and land-use/land-cover mapping. Suggestions on environmental governance can be provided by the remote sensing technology.

The experiment was conducted in Xuzhou of Jiangsu Province. Xuzhou is an important coal producing area in China, which has confirmed reserves of more than 3.9 billion tons with an annual output of about 25 million tons. However, the coal mining area may lead to surface subsidence and soil quality declining, threatening the safety of residence and crop cultivation. Meanwhile, it can induce secondary geological disasters. Since the tailing reservoirs are stacking with tailings of high acid and
alkali concentration, once it is destroyed by an extreme natural environment, it will cause flood plowing, damage roads, and other serious environmental pollution. The experimental area is located near the coal mining area, which has a security threat.

The experimental data were obtained by HySpex SWIR-384 and HySpex VNIR-1600 imaging spectrometer in Xuzhou on November 2014. The image obtained by HySpex VNIR-1600 with 0.19-m spatial resolution has 160 bands, which cover the spectral range of 415–992 nm, and the image obtained by HySpex SWIR-384 with 0.73-m spatial resolution has 288 bands, which cover the spectral range of 952–2508 nm. First, two images were preprocessed (including atmospheric correction and geometric correction). Second, the image obtained by HySpex VNIR-1600 was resampled to 0.73 m, and the two images were merged into one image after wiping off spectral overlap region and vapor absorption region. The new image has 436 bands and covers the spectral range of 415–2508 nm. Finally, an image with a size of 500 × 260 pixels was chosen as the experimental area. Based on field investigation, there are nine types of ground objects determined. The original image and ground truth image are shown in Fig. 15, and the number of labeled samples for each class is shown in Table IX. To evaluate the classification accuracy and verify the stability of the model, 10% and 5% of the labeled sample were selected as training samples separately, and rest for testing. According to the classification result by the parallel GBRBM, the best classification accuracy of the framework is using four layers (50 + 100 + 200 + 400). The accuracy comparison is shown in Tables X and XI, and classification maps are illustrated in Figs. 16 and 17. The results are similar to Section III-C; the GBRBM in parallel achieved the best accuracy and reducing the number of training samples also degraded the accuracy.

The classification maps show the land use in the experimental area, which can obtain the land use change by comparing the classification maps in different periods. From the current classification map, there are still large tracts of cultivated land around the tailing reservoirs, and the cultivated area has not been reduced by the existence of tailing. However, the potential risks still exist. The accumulation of tailings not only takes up a large of land resources but causes ecological pollution. It is mainly reflected in two aspects: one is dust pollution of surrounding land in the coal transportation, and the other is that heavy metals in tailing would pollute the cultivated land with groundwater. Therefore, it is helpful to find out the distribution of the tailing reservoirs for later recovery.
In this paper, a novel classifier based on parallel GBRBM was proposed, which was tested on three hyperspectral datasets. The experiments demonstrate that the parallel GBRBM is an effective feature extraction method. Compared to traditional classifiers and several deep learning models, the accuracy by parallel GBRBM-LR can improve about 2% to 5%. Although the parallel GBRBM-LR framework has a longer training time, its shorter prediction time makes it more suitable for a wide range of hyperspectral data. There are two significant parameters that impact the model classification accuracy: the number of layers in parallel and the number of hidden neurons, which can only be determined by experiments. In the future study, to make full use of spatial and spectral information, feature fusion will be introduced into this framework.

ACKNOWLEDGMENT

The authors would like to thank Prof. D. Landgrebe and P. Gamba for providing the AVIRIS and ROSIS datasets used in the experiments. The authors would also like to thank the Jiangsu Innovation Teams, CUMT.

REFERENCES


Kun Tan (SM’16) received the B.S. degree in information and computer science from Hunan Normal University, Hunan, China, in 2004, and the Ph.D. degree in photogrammetric and remote sensing from the China University of Mining and Technology, Jiangsu, China, in 2010. From September 2008 to September 2009, he was working toward the Joint Ph.D. degree in remote sensing at Columbia University, New York, NY, USA. From 2010 to 2018, he was with the Department of Surveying, Mapping, and Geoinformation, China University of Mining and Technology, Xuzhou, China. He is currently a Professor with East China Normal University, Shanghai, China. His research interests include hyperspectral image classification and detection, spectral unmixing, quantitative inversion of land surface parameters, and urban remote sensing.

Fuyu Wu was born in Shandong, China, in 1993. He received the B.S. degree in remote sensing science and technology from Wuhan University, Wuhan, China, in 2015. He is currently working toward the Ph.D. degree in photogrammetry and remote sensing at the China University of Mining and Technology, Xuzhou, China. His research interests include the hyperspectral imagery processing and ecological monitoring.

Qian Du (S’08–M’00–SM’05–F’18) received the Ph.D. degree in electrical engineering from the University of Maryland—Baltimore City, Baltimore, MD, USA, in 2000. She is currently a Bobby Shackoul Professor with the Department of Electrical and Computer Engineering, Mississippi State University, Starkville, MS, USA. Her research interests include the hyperspectral remote sensing image analysis, pattern recognition, and machine learning. Dr. Du was the Co-Chair for the Data Fusion Technical Committee of the IEEE Geoscience and Remote Sensing Society from 2009 to 2013, and the Chair for Remote Sensing and Mapping Technical Committee of the International Association for Pattern Recognition from 2010 to 2014. She is currently the Chief Editor for the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING. She is a fellow of SPIE—International Society for Optics and Photonics.

Peijun Du (M’07–SM’12) received the Ph.D. degree in photogrammetric and remote sensing from China University of Mining and Technology, Jiangsu, China, in 2001. He is currently a Professor of photogrammetry and remote sensing with the Department of Geographic Information Sciences, Nanjing University, Nanjing, China, and the Deputy Director of the Key Laboratory for Satellite Surveying Technology and Applications with the National Administration of Surveying and Geoinformation, China. His research interests include remote sensing image processing and pattern recognition; remote sensing applications; hyperspectral remote sensing information processing; multi-source geospatial information fusion and spatial data handling; integration and applications of geospatial information technologies; and environmental information science. Prof. Du is an Associate Editor of the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS.

Yu Chen received the master’s degree in photogrammetry and remote sensing from the China University of Mining and Technology (CUMT), Xuzhou, China, in 2012, and the Ph.D. degree in earth and planetary science from the University of Toulouse, Toulouse, France, in 2017. In 2013, she was an Assistant Researcher with the Géosciences Environment Toulouse Laboratory, French National Center for Scientific Research. She is currently a Lecturer with CUMT. Her current research focuses on SAR interferometry, with particular emphasis on its application for geophysical studies.