A Hyperspectral Image Classification Method Based on Weight Wavelet Kernel Joint Sparse Representation Ensemble and β-Whale Optimization Algorithm

Mingwei Wang , Zitong Jia, Jianwei Luo , Maolin Chen , Shuping Wang, and Zhiwei Ye

Abstract—Joint sparse representation (JSR) is a commonly used classifier that recognizes different objects with core features extracted from images. However, the generalization ability is weak for the traditional linear kernel, and the objects with similar feature values associated with different categories are not sufficiently distinguished especially for a hyperspectral image (HSI). In this article, an HSI classification technique based on the weight wavelet kernel JSR ensemble model and the β-whale optimization algorithm is proposed to conduct pixel-level classification, where the wavelet function is acted as the kernel of JSR. Moreover, ensemble learning is used to determine the category label of each sample by comprehensive decision of some subclassifiers, and the β function is utilized to enhance the exploration phase of the whale optimization algorithm and obtain the optimal weight of subclassifiers. Experimental results indicate that the performance of the proposed HSI classification method is better than that of other newly proposed and corresponding approaches, the misclassification and classified noise are eliminated to some extent, and the overall classification accuracy reaches 95% for all HSIs.

Index Terms—β function, ensemble learning, hyperspectral image (HSI) classification, joint sparse representation (JSR), wavelet kernel, weight setting.

I. Introduction

IN RECENT years, hyperspectral remote sensing sensors have been applied to collect images with enough spectral

Mingwei Wang and Zitong Jia are with the Institute of Geological Survey, China University of Geosciences, Wuhan 430074, China (e-mail: wangmingwei@cug.edu.cn; jiazitong.1125@foxmail.com).

Jianwei Luo and Shuping Wang are with the Hubei Cancer Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430079, China (e-mail: wuljw@126.com; wsping52@126.com).

Maolin Chen is with the School of Civil Engineering, Chongqing Jiaotong University, Chongqing 400074, China (e-mail: maolinchen@whu.edu.cn).

Zhiwei Ye is with the School of Computer Science, Hubei University of Technology, Wuhan 430068, China (e-mail: weizhiye121@163.com).

Digital Object Identifier 10.1109/JSTARS.2021.3056198

resolution that contains hundreds of bands and allows the discrimination of objects with similar attributes [1]. A hyperspectral image (HSI) has been considered as an applicable tool for Earth observation because of its ability to obtain independent and continuous bands, analyze information from visible to near-infrared wavelength ranges, and supply multiple features from the fixed wavelength. It provides abundant spectral information and has a huge potential for the interpretation of different ground objects [2], [3]. As a result, the analysis of HSI has become a subject of research interest in remote sensing, which has been applied in a series of fields such as quantitative analysis [4], environmental monitoring [5], and land-cover mapping [6]. In addition, image classification is a significant step in identifying object types on the Earth's surface, and HSI classification aims to distinguish each sample into a discrete group of specific category labels [7], [8].

Existing HSI classification techniques are separated into two scopes: unsupervised and supervised [9]. For unsupervised techniques, fuzzy clustering [10], rough set [11], and iterative self-organizing data analysis technique algorithm [12] have been utilized to classify HSI samples. In these techniques, the process of classification is only based on the characteristics of feature values, and the misclassification is obvious as spectral characteristics are similar for different objects. For supervised techniques, active learning [13], random forest [14], and support vector machine (SVM) [15] have been utilized to obtain the category label of each pixel. Although these classifiers make full use of spectrum difference, the category label of the current pixel is usually impacted by the feature values on the neighbor. Therefore, several ideas are presented to synthesize the spatial and spectral characteristics of HSIs, and they are based on the hypothesis that samples within a local space have approximate spectral characteristics and express the same objects [16], [17]. In addition, HSI classification based on a deep learning model has been proposed to sufficiently synthesize spatial and spectral information, thus obtaining the category label of each pixel, but it is supported by the sufficient number of training samples and the sufficient amount of iterations, which is time-consuming as the data dimension increases [18].

As a well-behaved supervised classification model, sparse representation (SR) is used to recover the original data and

report class discriminative information, which has been widely used in the field of pattern recognition [19]. In addition, joint representation is presented to promote the stability of SR and boost its capability [20]. For HSI classification, samples with the same category are theoretically located in a low-dimensional subspace, and joint SR (JSR) makes associative decision on neighbor pixels as to which are feasible and particularly suitable for HSIs [21]. For the classification process, a testing sample is similarly expressed by a certain number of rules from the training dictionary, and the reconstructed matrix is utilized to determine the category label by searching for the minimum [22]. For instance, Peng *et al.* [23] designed a local adaptive JSR (LAJSR) technique for HSI classification; the dictionary construction and SR phases were improved by choosing representative rules from an additional dictionary. Tu *et al.* [24] proposed an HSI classification approach based on the balance of JSR and correlation coefficient (JSR-CC), which synthetically considered both local spatial and spectral similarities. Furthermore, the reconstructed matrix is usually computed by a linear kernel, making it difficult to reflect the inner product of nonlinear mapping between input spectral features and output category labels. Hence, Zhang *et al.* [25] proposed a novel HSI classification technique using JSR and nonlinear kernel extension, which mapped the input into a high-dimensional space to separate different objects and reflect better performance than that using the linear kernel. However, the category label is determined by kernel computing of higher order polynomial; the misclassification for specific categories will be enlarged if the order is uncertain within effective time.

The category label is obtained by the probability of kernel computing for JSR, and it is the same as other nonlinear classifiers in mechanism, such as k-nearest neighbor (KNN) and SVM. Moreover, the wavelet function is a series of formulations that are based on wavelet analysis and adequately keeps regularity and orthogonality; it has been employed in the field of HSI classification as the kernel of KNN and SVM to substitute for a linear kernel [26], [27]. As a result, the wavelet function is able to act as the kernel of JSR in theory. Ensemble learning is a machine learning paradigm that synthesizes multiple subclassifies to solve the same problem; better discrimination ability is obtained than the single classifier according to different emphases of subclassifiers especially for indeterminate objects and has been applied for HSI classification [28], [29]. However, the category label is usually obtained by the voting strategy for ensemble learning; the discrimination is confused if the votes are similar for two categories. As for JSR, the category label is assigned by searching for the minimum of reconstructed error for each sample, and the reconstructed matrix of ensemble learning can be updated by that of subclassifiers with weight setting. A higher weight means that the subclassifier produces more contribution for classification, and a suitable weight setting is able to balance the reconstructed error of subclassifiers [30]. In general, how to obtain the optimal weight of subclassifiers is seen as a combination optimization problem, and it can be solved by the swarm intelligence algorithm with heuristic search guiding strategies [31]. Among them, the whale optimization algorithm (WOA) is a newly proposed swarm intelligence algorithm and

has been widely used in diverse applications especially for weight optimization [32], [33]. However, the convergence rate is not fast enough with a fixed population updating equation and the small probability of local search. Nowadays, the factorial function with a single parameter has been combined with the swarm intelligence algorithm to enhance the exploration phase, but it is not adapted to various population updating conditions such as WOA with multiple parameters [34], [35]. Here, the β function is combined with the WOA, two parameters are corresponding to two evolution processes, and the weight is adaptively located on the range of [0,1].

Therefore, an HSI classification technique based on the weight wavelet kernel JSR ensemble (W^2 JSRE) model and the β-WOA is proposed to conduct pixel-level classification for HSIs. Because the spectral feature is output by 16 bits, the discrimination is not significant for different categories, and the misclassification is obvious as the dataset is mapped into the linear kernel. The classification accuracy is improved as the wavelet function is acted as the kernel of KNN and SVM; the dataset is mapped into quadratic, exponential, and trigonometric functions with different types and has been utilized in the field of HSI classification, but it is not acted as the kernel of JSR for previous work. In addition, a series of subclassifiers based on JSR with wavelet kernels are integrated by ensemble learning, the wavelet kernel of JSR concerns on the homogeneity for each subclassifier, and the ensemble with multiple wavelet kernels emphasizes the heterogeneity. Furthermore, the swarm intelligence algorithm is widely used to solve the nonpolynomial hard problem, such as weight optimization, and the β-WOA is designed to obtain the optimal weight of subclassifiers, and the category label is output by total reconstructed error minimization of ensemble learning. The main contributions of this article are concluded as follows.

1) To improve the scale of mapping, the wavelet function is acted as the kernel of JSR, and the HSI dataset is mapped into quadratic, exponential, and trigonometric functions with different types.
2) To synthesize the homogeneity and heterogeneity of the JSR ensemble, the W^2 JSRE model is proposed by using different types of wavelet function as the kernel, and the classification map is output by pixel level.
3) To balance the reconstructed error of subclassifiers, weight setting is conducted for ensemble learning, and the category label is obtained by total reconstructed error minimization.
4) To enhance the exploration phase of the WOA, the β-WOA is designed by fusing the β function into two evolution processes of the WOA, and the optimal weight of subclassifiers is obtained.

The overall construction of this article is listed as follows. Section II describes the related work of JSR and WOA. Section III illustrates the principle of the proposed W^2 JSRE model and β-WOA and the fundamental process of HSI classification. Section IV analyzes the experimental results and expends discussion of data statistics and visual senses. Finally, Section V concludes this article.

II. RELATED WORK

A. Basic Theory of JSR

JSR is devoted to minimizing the reconstructed error of some independent SRs, and the inner correlations between different SRs are synthetically considered. In the HSI, spectral characteristics of a pixel are strongly correlated with its neighbor pixels, which means that they belong to the same object with large probability, and the spatial correlations are ensured by supposing that neighbor pixels within a local space are jointly indicated by some common-sense rules from a training dictionary [36]. In particular, the size of local space at center pixel y_t is signed by $l \times l$, and pixels within such a space are marked by y_i, where $i = 1, 2, \ldots, l \times l$. All of the above pixels are stacked into a matrix $Y = [y_1, y_2, \ldots, y_t, \ldots, y_{l \times l}] \in R^{b \times l^2}$. The matrix is succinctly represented as follows:

$$Y = [y_1, y_2, \ldots y_t, \ldots, y_{l \times l}] = [D\alpha_1, D\alpha_2, \ldots D\alpha_t, \ldots, D\alpha_{l \times l}]$$
$$= D[\alpha_1, \alpha_2, \ldots \alpha_t, \ldots, \alpha_{l \times l}] = DA \qquad (1)$$

where $A = [\alpha_1, \alpha_2, \ldots, \alpha_t, \ldots, \alpha_{l \times l}] \in R^{b \times l^2}$ is the recovered data with regard to Y. The selected rules in D are assigned by rows and columns of elements that are not equal to 0 in $[\alpha_1, \alpha_2, \ldots, \alpha_t, \ldots, \alpha_{l \times l}]$, by setting part of rows as the value of 0 on the reconstructed matrix A. The neighbor pixel Y is expressed by a subset of common-sense rules. Afterward, the matrix is recovered by seeking the equation to represent the following optimization problem:

$$\hat{A} = \arg \min_A \|Y - DA\|_F \quad \text{s.t.} \|A\|_{\text{row},0} \leq K \qquad (2)$$

where $\|A\|_{\text{row},0}$ is the joint sparse norm that finds the most representative nonzero rows in A, and $|\cdot|_F$ is the Frobenius norm. As $\hat{A}$ is recovered, the category label at the center pixel y_t is judged by the reconstructed error that is defined as follows:

$$\text{label}(y_t) = \arg \min r(y) = \arg \min_{i=1,2,\ldots,c} \|Y - D_i \hat{A}_i\|_2 \qquad (3)$$

where $\hat{A}_i$ indicates the rows in $\hat{A}$ associated with the category index of i.

B. Mathematical Model of WOA

In 2016, Mirjalili designed a swarm intelligence algorithm called WOA that is based on the predatory strategy of humpback whales. Humpback whales tend to catch crowd of krill or small fishes near the surface. The process is conducted by producing specific bubbles with a ring path, and the operator is separated into three parts: encircling prey, spiral bubble-net attacking, and searching for prey. The main procedure for the WOA is depicted as follows [37]:

Encircling prey: Humpback whales have the ability to search for the position of prey and surround them, and the mechanism of global search is represented by the process. It is assumed that the position of optimal solution is the objective prey or it is the proximate solution moving close to the optimum in theory, and others should endeavor to motivate their positions toward to it.

The process is written as follows:

$$\vec{S} = |\vec{C} \cdot X^*(t) - X(t)| \qquad (4)$$
$$X(t+1) = X^*(t) - \vec{A} \cdot \vec{S} \qquad (5)$$

where t is the number of current iterations, $X^*(t)$ is the position of prey, and $X(t)$ and $X(t+1)$, respectively, represent the position of humpback whales in the current and the next procedure. $\vec{A}$ and $\vec{C}$ are the variable vectors that are expressed as $\vec{A} = 2\vec{a} \cdot \vec{r} - \vec{a}$ and $\vec{C} = 2 \cdot \vec{r}$, $\vec{a} = 2 - 2 * t/T$ is gradually decreased within the scope of [2,0], T is the maximum number of iteration, and $\vec{r}$ is a random number on the range of [0,1].

Bubble-net attacking: Each humpback whale moves close to the prey within a compact ring and follows a spiral-shaped path in the meantime, and the mechanism of local search is represented by the process. A probability of 0.5 is set to choose whether following the compact ring or spiral mechanism, the position of humpback whale is renewed. The formulation of the process is expressed as follows:

$$X(t+1) = \begin{cases} X^*(t) - \vec{A} \cdot \vec{S}, & \text{if} \quad p < 0.5 \\ \vec{S'} \cdot e^{bl} \cdot \cos(2\pi l) + X^*(t), & \text{if} \quad p \geq 0.5 \end{cases} \qquad (6)$$

where $\vec{S'}$ is the distance of current humpback whale to prey, which is expressed as $\vec{S'} = |X^*(t) - X(t)|$, $b = 1$ represents a constant number that is the situation of logarithmic spiral, and l and p are two random numbers, respectively, within the scope of [−1,1] and [0,1].

Searching for prey: The position of current humpback whale is updated according to the random walk strategy rather than the best humpback whale, the strategy of random search is reflected by the process, and the details are expressed as follows:

$$\vec{S} = |\vec{C} \cdot X_{\text{rand}} - X(t)| \qquad (7)$$
$$X(t+1) = X_{\text{rand}} - \vec{A} \cdot \vec{S} \qquad (8)$$

where X_{rand} indicates the position of a random humpback whale selected from the population.

III. PROPOSED METHODOLOGY

A. Classification Process With W^2 JSRE

As for (3), the category label is determined by reconstructed error minimization, and it is computed on the same scale with the linear kernel, which makes it difficult to express the difference of feature values on multiple scales and emerge the relationship of nonlinear mapping in detail. The basic theory of wavelet analysis is to combine wavelet basis that builds an arbitrary function following the time series t, there are five types of wavelet function that are proposed by analytical expressions with compactly supported and can be decomposed to different scales, and they are defined as follows [38], [39]:

$$f_1(t) = \exp(-t^2/2) \qquad (9)$$
$$f_2(t) = (1 - t^2) \cdot \exp(-t^2/2) \qquad (10)$$
$$f_3(t) = \cos(1.75 \cdot t) \cdot \exp(-t^2/2) \qquad (11)$$

$$f_4(t) = \frac{\sin(0.5\pi \cdot t)}{0.5\pi \cdot t} \cdot \cos(1.5\pi \cdot t) \tag{12}$$

$$f_5(t) = \frac{e^{i4\pi \cdot t} - e^{i2\pi \cdot t}}{i2\pi \cdot t}. \tag{13}$$

The wavelet function contents the fixed condition of shift-invariant form, it is based on the inner product of nonlinear mapping on different scales, and the difference between the original and recovered data can be represented by shift-invariant form [41]. Nowadays, the wavelet function is acted as the kernel of wavelet kernel SVM (WSVM) and wavelet kernel KNN (WKNN), and the classification result is improved as the dataset is mapped into different scales. More importantly, the dataset with ten thousands of samples is difficultly expressed by a linear kernel mapping. For JSR, the learning mechanism is the same with SVM and KNN, and the wavelet function can be acted as the kernel of JSR; the reconstructed error is defined on the basis of (3) and is expressed as follows:

$$r_1(y) = \exp(-\|Y - D_i\hat{A}_i\|_2/2) \tag{14}$$

$$r_2(y) = (1 - \|Y - D_i\hat{A}_i\|_2) \cdot \exp(-\|Y - D_i\hat{A}_i\|_2/2) \tag{15}$$

$$r_3(y) = \cos(1.75 \times \|Y - D_i\hat{A}_i\|_1) \cdot \exp(-\|Y - D_i\hat{A}_i\|_2/2) \tag{16}$$

$$r_4(y) = \frac{\sin(0.5\pi \times \|Y - D_i\hat{A}_i\|_1)}{0.5\pi \times \|Y - D_i\hat{A}_i\|_1} \cdot \cos(1.5\pi \times \|Y - D_i\hat{A}_i\|_1) \tag{17}$$

$$r_5(y) = \frac{e^{i4\pi\|Y - D_i\hat{A}_i\|_1} - e^{i2\pi\|Y - D_i\hat{A}_i\|_1}}{i2\pi\|Y - D_i\hat{A}_i\|_1} \tag{18}$$

where "·" represents the inner product between the vectors of reconstructed error with two different scales, and the original dataset is mapped into quadratic, exponential, and trigonometric functions with different types. Experimental results demonstrate that a scale parameter is involved in the dilation and, thus, can be naturally used to accommodate the multiscale phenomenon [40].

The category label of a sample is determined by five subclassifiers (JSRs) with different wavelet kernels at the same time, which is able to improve the discrimination ability compared with single JSR and linear kernel. The significance of subclassifiers is decided by weight setting, and the reconstructed error of the proposed W^2 JSRE model is computed as follows:

$$\text{label}(y_t) = \arg\min \sum_{j=1}^{5} \omega_j \times r_j(y) \tag{19}$$

where ω_j is the weight of the jth subclassifier, which is directly multiplied with the reconstructed matrix, and weight represents the significance of subclassifiers. It is seen as a fuzzy quantitative analysis for the ensemble learning of JSRs, and the performance is better than the traditional voting strategy with fixed category analysis.

B. Weight Optimization With β-WOA

The exploration phase is represented by searching for prey to conduct random walk, which is computed by the position

of a random humpback whale, but the operation efficiency is decreased by random number generation and the evolution trend is uncollected for the enlarge of $\vec{S}$ in (8). The β function is a factorial function with analytic continuation in the complex plane; two parameters γ and η are defined to adjust the value. For the improvement of the swarm intelligence algorithm, it is necessary to weaken the random process and synthesize multiple parameters updating the individuals. The value range of the β function is [0,1], which is adapted to the weight ω_j of subclassifiers. As for the proposed β-WOA, the exploration phase is based on the β function instead of searching for prey, and, respectively, acting on encircling prey and bubble-net attacking, which is defined as follows:

$$X(t+1) = \int_0^1 t^{\gamma-1}(1-t)^{\eta-1}dt \tag{20}$$

where

$$\gamma = (X^*(t) - \vec{A} \cdot \vec{S})^{-1}$$
$$\eta = (\vec{S}' \cdot e^{bl} \cdot \cos(2\pi l) + X^*(t))^{-1}. \tag{21}$$

There is no random humpback whale that needs to be extracted, all of individuals are, respectively, computed by two processes of encircling prey and bubble-net attacking, they are corresponding to γ and η of β function, and the population is updated by (20) afterward. As a result, the global and local processes are integrated for each individual and iteration, and time complexity is decreased by no random sample generation. Moreover, the coding length of the β-WOA is equal to 5, which is the same as the number of subclassifiers, and directly represents the weight of subclassifiers.

C. Definition of the Objective Function

The key issue of HSI classification based on the W^2 JSRE model is how to establish a reasonable mapping between the solution and the β-WOA. As for weight setting, it is expressed by a constant on the range of [0,1] for subclassifiers and corresponding to a bit of β-WOA. Each individual of β-WOA includes 5 bits: the first bit represents the weight of the first JSR (subclassifier), the second bit is the weight of the second JSR (subclassifier), and so on. The entire code indicates the solution about the optimal weight of the W^2 JSRE model, and the fitness value is computed according to the average entropy of the reconstructed matrix, which is defined as follows:

$$F(i) = -\sum_{i=1}^{s} \min_j \hat{A}_{ij}\log_2(\hat{A}_{ij})/s \tag{22}$$

where s is the scale of testing samples, and j is the category index that takes on the minimum for the ith testing sample. A larger fitness value means that the reconstructed error is smaller, and the category label is more likely to obey the true distribution.

D. Implementation of the Proposed Method

The proposed HSI classification technique is easy to be fulfilled. The W^2 JSRE model is used for pixel-level classification of HSIs and the category label is obtained for each sample, the

---

Algorithm 1: HSI Classification Based on the W^2 JSRE Model and the β-WOA.

Input: Construct the training samples for each category from original HSIs, and the iteration number of β-WOA is $t = 0$.

Output: The classification maps based on pixel level for all HSIs.

1: Input HSIs, and transform it to the reasonable format that is directly read by ENVI software;

2: Randomly choose 10% of pixels for each category that are acted as training samples;

3: Build the W^2 JSRE model, and assign the kernel of subclassifiers according to five different wavelet functions;

4: Generate initial population of β-WOA, and express the weight of subclassifiers;

5: **while** The algorithm does not reach the termination condition **do**

6:    $t = t + 1$;

7:    Conduct testing for the W^2 JSRE model, and compute the fitness value of each humpback whale by (22);

8:    Employ encircling prey and bubble-net attacking for the whole population;

9:    Compute the parameters γ and η of the β function;

10:    Use the β function instead of searching for prey with exploration phase;

11:    **if** The fitness value is higher than last iteration **then**

12:       Replace the position of current best humpback whale;

13:    **end if**

14: **end while**

15: Output the global optimal solution of the β-WOA, and the optimal weight of subclassifiers;

16: **return** The category label of each pixel according to the minimum of reconstructed error.

β-WOA is used to search for the optimal weight of subclassifiers (JSRs), and the exact flow is listed as follows.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

A. Data Description

To evaluate the performance of the proposed HSI classification technique based on the W^2 JSRE model and the β-WOA, three public collected HSIs and two measured airborne HSIs are used in the experiments.

The first HSI was acquired by the ROSIS sensor during a flight campaign over Pavia University, Italy, and the geometric resolution was 1.3 m [42]. The image was composed of 610×340 pixels with 103 spectral bands. Fig. 1 displays the ground truth of PaviaU scene. The number and names of corresponding categories that were used are shown in Table I.

The second HSI was collected by the AVIRIS sensor and covered the agricultural region of Indian Pines, India, in 1992 [42]. The spectral range was 0.4–2.5 μm with a spectral resolution

Fig. 1. Original image and reference map of PaviaU.

TABLE I
LAND-COVER CLASSES IN PAVIAU IMAGE

Class number	Class name	No. of labelled samples
1	Asphalt	6631
2	Meadows	18,649
3	Gravel	2099
4	Trees	3064
5	Painted metal sheets	1345
6	Bare Soil	5029
7	Bitumen	1330
8	Self-Blocking Bricks	3682
9	Shadows	947
	Total	42,776

Fig. 2. Original image and reference map of Indian.

about 10 nm, and the image was composed of 145×145 pixels and 220 spectral bands with a spatial resolution of 20 m. Fig. 2 displays the ground truth of Indian scene. The number and names of corresponding categories that were used are shown in Table II.

The third HSI was collected by the 224-band AVIRIS sensor over Salinas Valley, California, and it was characterized by high spatial resolution. The image was composed of 512×217 pixels and available only as sensor radiance data, and 20 water absorption bands were discarded [42]. Fig. 3 displays the ground

TABLE II
LAND-COVER CLASSES IN INDIAN IMAGE

Class number	Class name	No. of labelled samples
1	Alfalfa	46
2	Corn-notill	1428
3	Corn-min	830
4	Corn	237
5	Grass/Pasture	483
6	Grass/Trees	730
7	Grass/Pasture-mowed	28
8	Way-windrowed	478
9	Oats	20
10	Soybeans-notill	972
11	Soybeans-min	2455
12	Soybean-clean	593
13	Wheat	205
14	Woods	1265
15	Bldg-Grass-Tree-Drives	386
16	Stone-steel towers	93
	Total	10,249

TABLE III
LAND-COVER CLASSES IN SALINAS IMAGE

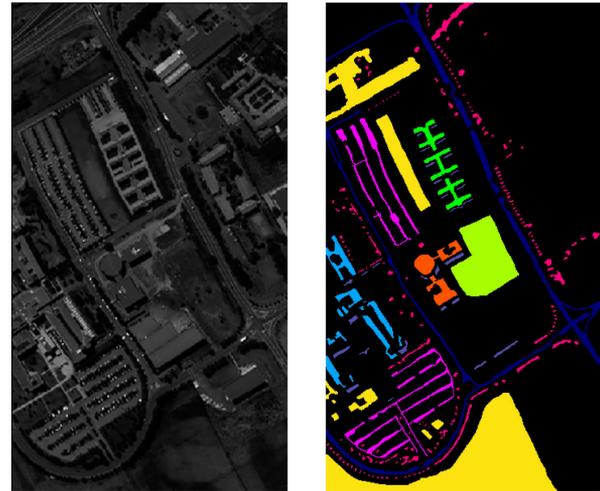| Class number	Class name	No. of labelled samples
1	Brocoli_green_weeds_1	2009
2	Brocoli_green_weeds_2	3726
3	Fallow	1976
4	Fallow_rough_plow	1394
5	Fallow_smooth	2678
6	Stubble	3959
7	Celery	3579
8	Grapes_untrained	11,271
9	Soil_vinyard_develop	6203
10	Corn_senesced_green_weeds	3278
11	Lettuce_romaine_4wk	1068
12	Lettuce_romaine_5wk	1927
13	Lettuce_romaine_6wk	916
14	Lettuce_romaine_7wk	1070
15	Vinyard_untrained	7268
16	Vinyard_vertical_trellis	1807
	Total	54,129

Fig. 4.   Original image and reference map of XionganS.

TABLE IV
LAND-COVER CLASSES IN XIONGANS IMAGE

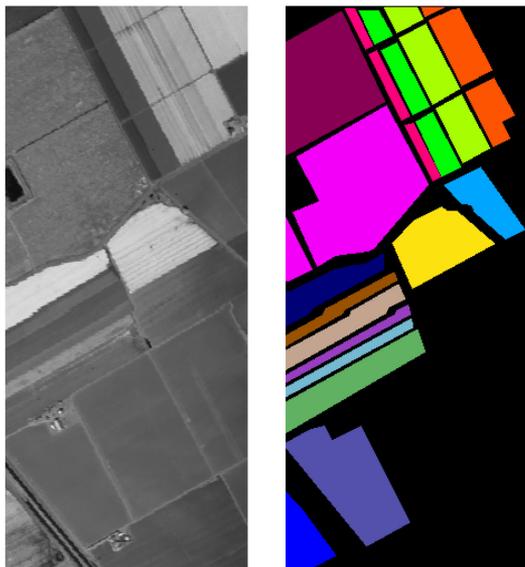| Class number	Class name	No. of labelled samples
1	Water	3877
2	Building	6568
3	Vegetation	3143
4	Field	7147
5	Bare land	2623
6	Path	843
7	Grass	6199
	Total	30,400

Fig. 3.   Original image and reference map of Salinas.

truth of Salinas scene. The number and names of corresponding categories that were used are shown in Table III.

The fourth HSI was collected by the CASI sensor over the suburban area of Xiongan, China, in the summer of 2017. The spectral range was 0.36–1.05 μm with a spectral resolution of 7.2 nm, and the image was composed of 160×190 pixels with 96 spectral bands. Fig. 4 shows the ground truth of XionganS scene. The number and names of corresponding categories that were used are shown in Table IV.

The fifth HSI was acquired by the SASI sensor over the urban area of Xiongan, China, in the spring of 2018. The spectral range was 1.0–2.5 μm with a spectral resolution of 15 nm, and the image was composed of 270×232 pixels with 100 spectral bands. Fig. 5 shows the ground truth of XionganU scene. The
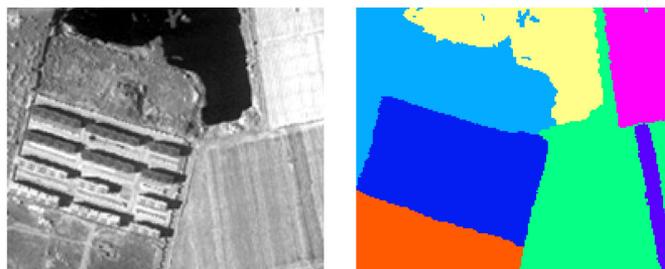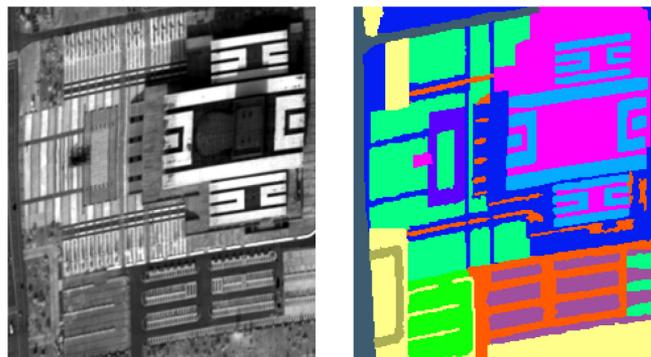

Fig. 5.   Original image and reference map of XionganU.

TABLE V
LAND-COVER CLASSES IN XIONGANU IMAGE

Class number	Class name	No. of labelled samples
1	Vegetation	6670
2	Bare land/Pitch	10,924
3	Marble	5492
4	Field	11,255
5	Bare land/Mud	9666
6	Shrub	1416
7	Building	5980
8	Grass	2771
9	Path	706
10	Fence	3964
11	Road	3796
	Total	62,640

TABLE VI
PARAMETERS SETTING OF DIFFERENT ALGORITHMS

Parameters	Value
Population size	10
Dimension	5
Number of iteration	30
c_1, c_2 Acceleration constants in PSO	2.0
f_m Mutation factor in DE	0.6
C_R Crossover rate in DE	0.9
p_a Detecting probability in CS	0.25
$\vec{a}$ Correlation coefficient in GWO	[2,0]
ω Stepping factor in ALO	[2,6]
$\vec{a}$ Correlation coefficient in WOA and β-WOA	[2,0]

number and names of corresponding categories that were used are shown in Table V.

B. Parameters Setting of Different Algorithms

As for the β-WOA, there is one parameter that needs to be set by the corresponding reference [32]. Moreover, some commonly used swarm intelligence algorithms are also assessed to conduct weight optimization. As the illustration in Section III, the β-WOA is utilized here, whereas particle swarm optimization (PSO) [43], differential evolution (DE) [44], cuckoo search (CS) [45], grey wolf optimizer (GWO) [46], ant lion optimizer (ALO) [47], and standard WOA are utilized to make intuitive comparisons. All of the above algorithms are ended as the of evaluations reaches 300. Thirty independent operations are conducted because of the randomness of initial population. Although the computational complexity is $O(n\log n)$ for the algorithms above [48], there is no random humpback whale that needs to be extracted for the β-WOA, and each bit will be adaptively located on [0,1] by the range of β function, which will cost less CPU time than that of the standard WOA. The parameters of these algorithms are set by constants and based on the empirical value of corresponding references, and they are listed in Table VI.

C. Experimental Results on Swarm Intelligence Algorithms

In this subsection, evaluation of training samples with the weight optimized by different swarm intelligence algorithms is investigated. For five HSIs in Section IV-A, 10% of pixels for each category are randomly extracted as the training samples to obtain weights of subclassifiers. Table VII shows the experimental results with different swarm intelligence algorithms, where Fiv and Std represent the average and standard deviation of fitness value, respectively, and Time is the CPU time after 30 independent operations.

As for the data in Table VII, the optimization ability of the WOA is obviously better than that of PSO, DE, CS, GWO, and ALO, and the fitness value is higher than 0.30 for the five datasets. In addition, the β function is operated for encircling prey and bubble-net attacking of the basic WOA, which is acted as the heuristic information of exploration phase. More importantly, the fitness value is further improved compared with the basic WOA, which illustrates that the reconstructed error remains in a small interval between the original and recovered datasets. With regard to the operating efficiency, the convergence speed of the WOA is better than that of other algorithms because of less multiplications, and there is no random humpback whale that needs to be extracted for the β-WOA, and the CPU time is further decreased to some extent. Meanwhile, the weights optimized by the β-WOA are suitably assigned for five subclassifiers; these are set as 0.2242, 0.1101, 0.6585, 0.2343, and 0.3887 for the Indian dataset, and all of subclassifiers have a certain contribution for training. However, the category label may focus on one or two subclassifiers by using other algorithms. The weight is greater than 0.9 for a subclassifier, and the performance of ensemble learning does not sufficiently play. In brief, the optimization ability of the β-WOA is superior, and the convergence speed is fast enough to obtain the satisfactory weight, which is applicable for the practical work of sample training about HSI classification.

D. Experimental Results About HSI Classification on Pixel Level

In this subsection, five HSIs, named PaviaU, Indian, Salinas, XionganS, and XionganU, are utilized to conduct pixel-level classification of HSIs and verify the performance of the W^2 JSRE model and the β-WOA. Moreover, some corresponding and newly proposed HSI classification techniques such as JSR [22], LAJSR [23], JSR-CC [24], wavelet kernel JSR (WJSR), WKNN [26], WSVM [27], and deep learning model, such as fully convolutional networks (FCN) [49], discriminative stacked autoencoder (DSAE) [50], are also used to make an overall comparison. In addition, the classification results with different percentages of training samples (Indian image is not operated because of less number of samples for Alfalfa and Oats categories) and three subclassifiers of ensemble learning are also exhibited to make a further verification; the experiments are not conducted for LAJSR, JSR-CC, and WJSR because of the correlation of JSR-based techniques. The classification maps of different techniques are listed in
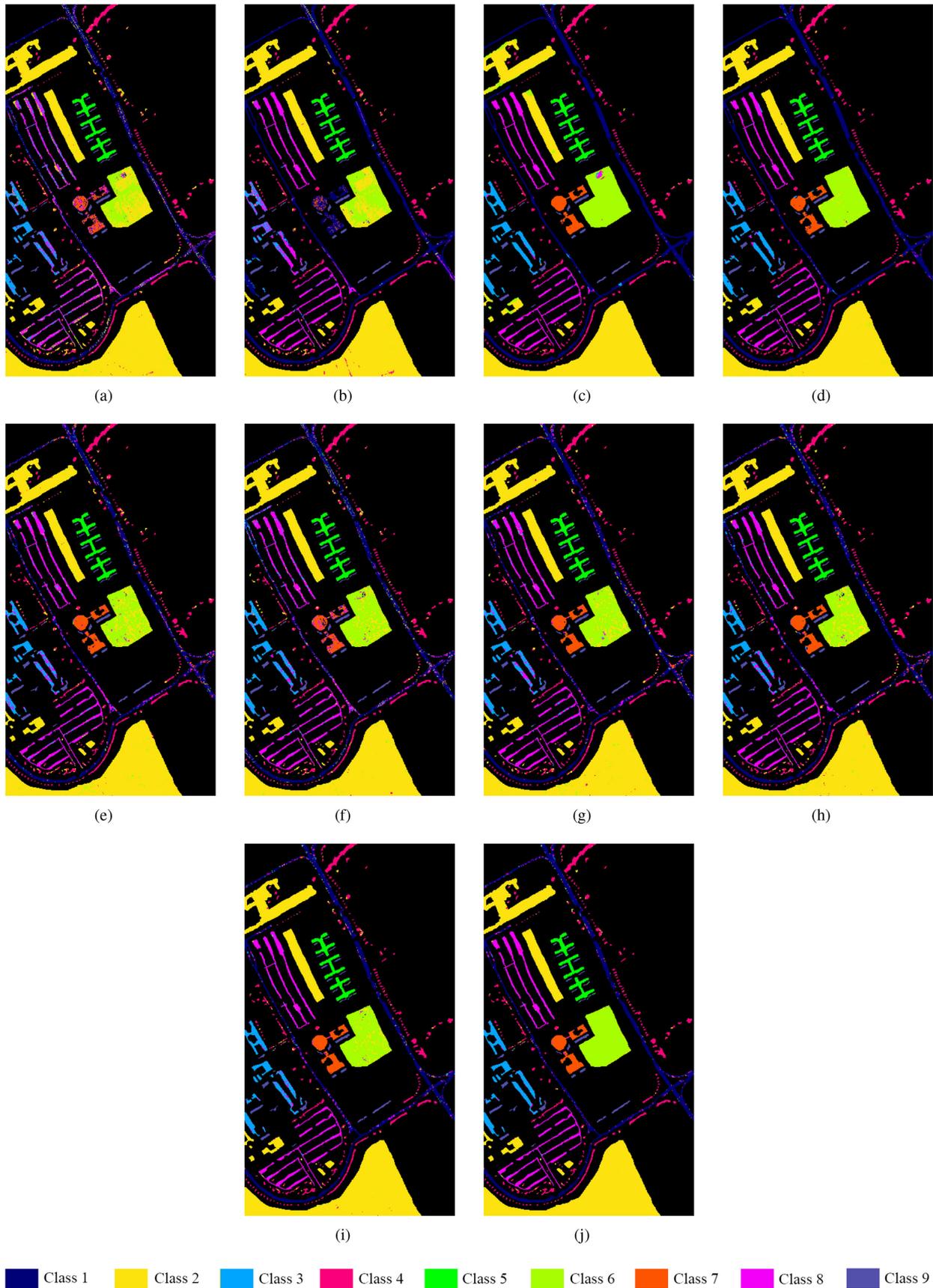
Fig. 6. Classification results of PaviaU image. (a) WKNN. (b) WSVM. (c) FCN. (d) DSAE. (e) JSR. (f) LAJSR. (g) JSR-CC. (h) WJSR. (i) W^2 JSRE (three subclassifiers). (j) W^2 JSRE (five subclassifiers).

TABLE VII
FITNESS VALUE AND CPU TIME FOR DIFFERENT ALGORITHMS

Dataset	Meas.	PSO	DE	CS	GWO	ALO	WOA	β-WOA
	Fiv	0.2973	0.2991	0.3002	0.3014	0.3028	0.3046	**0.3058**
PaviaU	Std	0.0054	0.0052	0.0048	0.0044	0.0041	0.0038	**0.0033**
	Time	0.5876	0.5931	0.6015	0.5887	0.6248	0.5862	**0.5828**
	Fiv	0.2925	0.2960	0.3039	0.3064	0.3082	0.3101	**0.3132**
Indian	Std	0.0093	0.0081	0.0076	0.0068	0.0062	0.0056	**0.0051**
	Time	0.2065	0.2102	0.2123	0.2073	0.2296	0.2058	**0.2037**
	Fiv	0.2959	0.3008	0.3056	0.3094	0.3118	0.3169	**0.3203**
Salinas	Std	0.0018	0.0014	0.0013	0.0010	0.0008	**0.0007**	**0.0007**
	Time	1.7072	1.7308	1.7637	1.7202	1.8173	1.7025	**1.6947**
	Fiv	0.2955	0.2967	0.2973	0.2984	0.2996	0.3010	**0.3019**
XionganS	Std	0.0117	0.0106	0.0102	0.0095	0.0078	0.0071	**0.0065**
	Time	0.3097	0.3133	0.3156	0.3109	0.3333	0.3073	**0.3049**
	Fiv	0.3078	0.3091	0.3101	0.3113	0.3119	0.3127	**0.3135**
XionganU	Std	0.0048	0.0043	0.0039	0.0036	0.0033	**0.0030**	0.0031
	Time	1.0527	1.0621	1.0754	1.0553	1.1207	1.0503	**1.0440**

Fig. 7. Classification results of Indian image. (a) WKNN. (b) WSVM. (c) FCN. (d) DSAE. (e) JSR. (f) LAJSR. (g) JSR-CC. (h) WJSR. (i) W^2 JSRE (three subclassifiers). (j) W^2 JSRE (five subclassifiers).

Fig. 8. Classification results of Salinas image. (a) WKNN. (b) WSVM. (c) FCN. (d) DSAE. (e) JSR. (f) LAJSR. (g) JSR-CC. (h) WJSR. (i) W^2 JSRE (three subclassifiers). (j) W^2 JSRE (five subclassifiers).
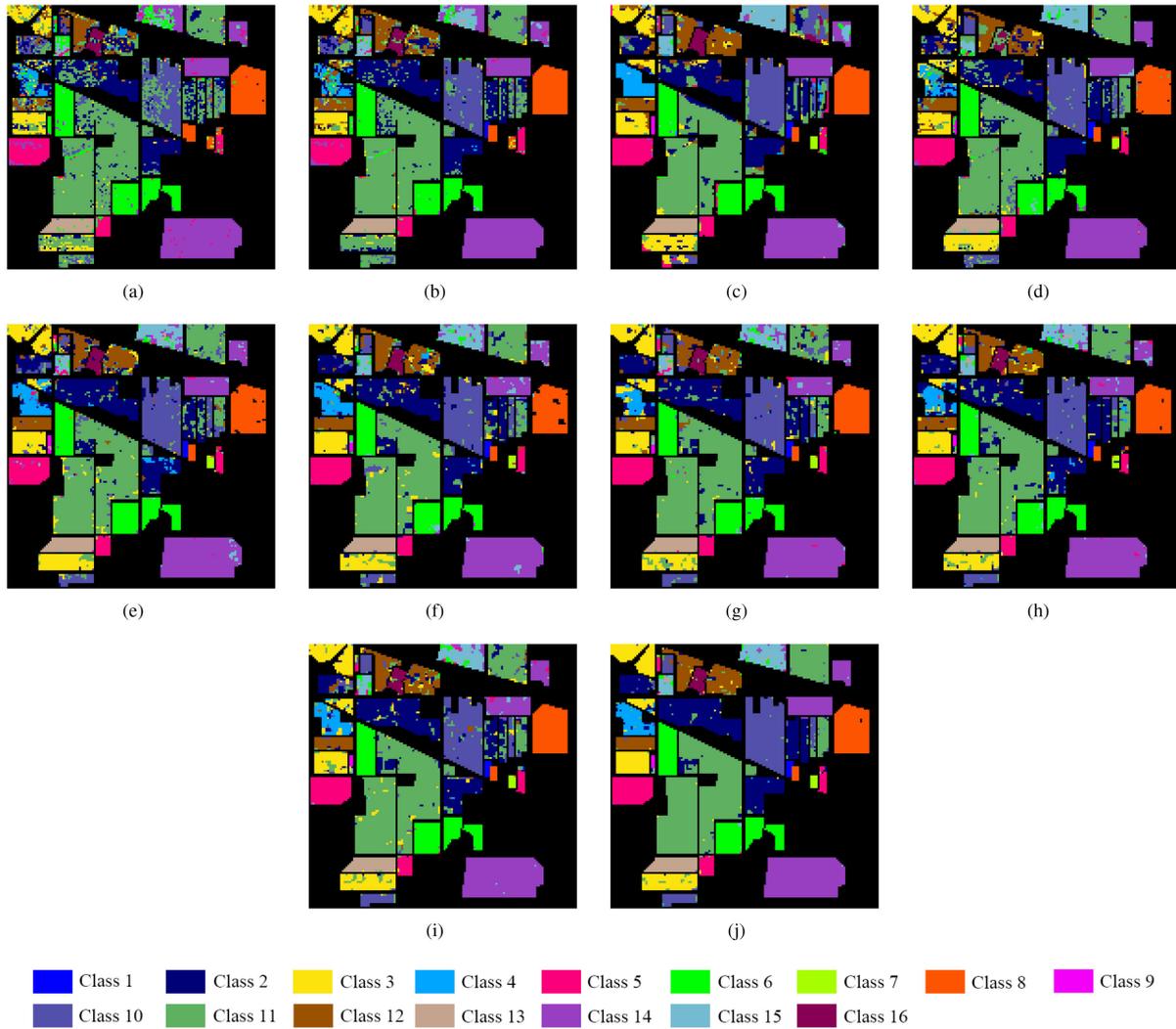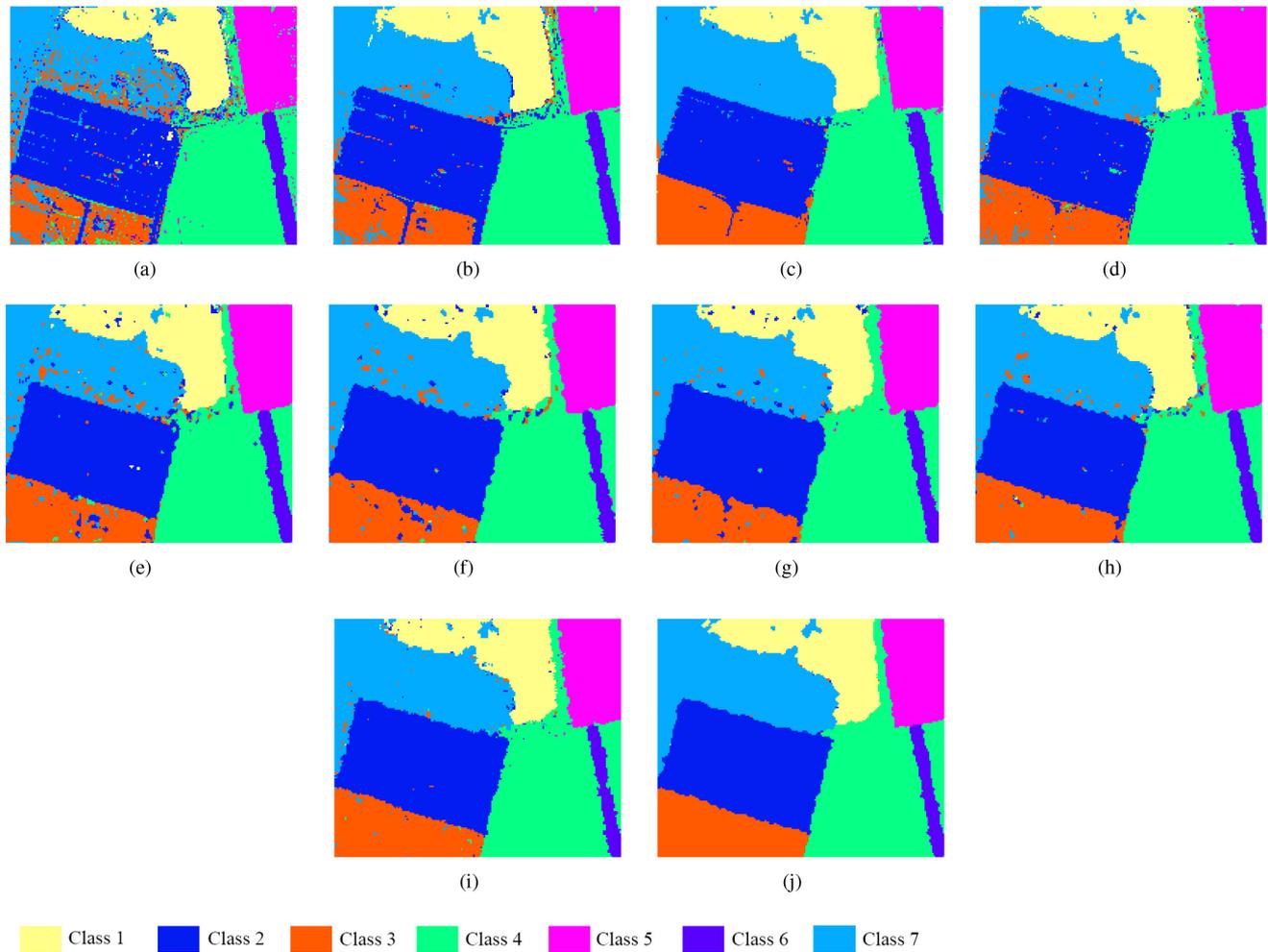
Fig. 9. Classification results of XionganS image. (a) WKNN. (b) WSVM. (c) FCN. (d) DSAE. (e) JSR. (f) LAJSR. (g) JSR-CC. (h) WJSR. (i) W^2 JSRE (three subclassifiers). (j) W^2 JSRE (five subclassifiers).
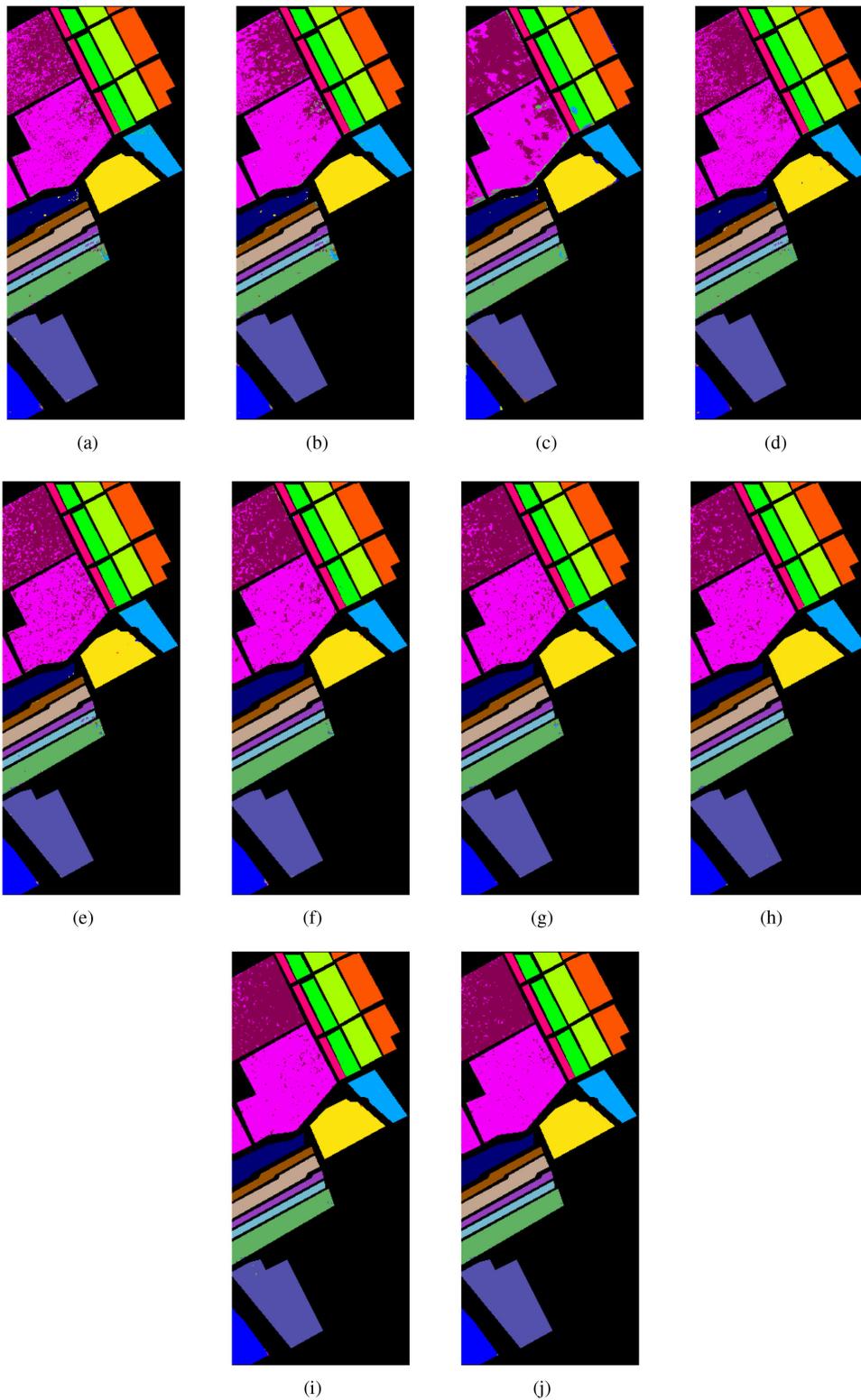
TABLE VIII
OA AND KAPPA COEFFICIENT FOR PAVIAU IMAGE

Class number Meas.	WKNN	WSVM	FCN	DSAE	JSR	LAJSR	JSR-CC	WJSR	W^{2}JSRE 3 sub-classifiers	W^{2}JSRE 5 sub-classifiers
1	77.5750	93.3494	97.6474	97.3458	80.8777	77.3337	83.9391	83.7129	87.1060	95.4306
2	99.0723	99.1903	96.7237	98.2895	98.5039	99.3351	98.8686	99.1689	99.7587	99.9196
3	74.5117	39.4474	97.2844	96.0457	79.0853	85.1358	82.9919	83.5636	89.0424	95.5693
4	82.5392	90.5352	97.5196	97.9112	88.8708	87.3042	90.8943	90.9269	91.9712	97.5522
5	99.3309	99.4052	99.9257	99.8513	99.7770	99.7026	100.0000	100.0000	100.0000	100.0000
6	51.4416	59.0177	97.9916	95.4862	74.8260	80.9704	82.8196	80.7914	88.3277	98.5086
7	60.3008	13.5338	97.5940	96.5414	93.3835	69.6992	94.7368	95.8647	96.1654	97.3684
8	72.1619	93.8077	96.7681	95.5731	90.9017	91.4992	93.2917	92.6670	95.7360	97.1754
9	96.7348	99.8944	99.8944	99.5776	95.5649	96.6209	96.6209	96.6209	96.7265	99.8944
OA(%)	96.7348	97.2994	99.4793	99.4740	98.0419	98.0535	98.4489	98.4195	98.8863	99.4990
Kappa	0.9089	0.9247	0.9855	0.9851	0.9455	0.9458	0.9568	0.9560	0.9595	0.9861
Time	150.8371	138.4107	3414.8419	1299.3674	154.0600	532.9002	726.6229	145.9274	233.4838	343.6807

Figs. 6–10, and Tables VIII–XII outline the overall classification accuracy (OA), Kappa coefficient, and CPU time of each HSI.

Based on the data in Tables VIII–XII, there are no samples that are accurately classified to Alfalfa or Oats categories for Indian image by using traditional techniques. The OA of JSR-based techniques is obviously better than that of WKNN and WSVM, and it is higher than 80% for all categories of XionganU and XionganS images. Compared with the linear kernel, the wavelet kernel improves the scale of mapping, and the Kappa coefficient has reached 0.91 for five images. As for the W^2 JSRE model, the OA is superior to 95% for five images, and the Kappa coefficient

Fig. 10. Classification results of XionganU image. (a) WKNN. (b) WSVM. (c) FCN. (d) DSAE. (e) JSR. (f) LAJSR. (g) JSR-CC. (h) WJSR. (i) W^2 JSRE (three subclassifiers). (j) W^2 JSRE (five subclassifiers).
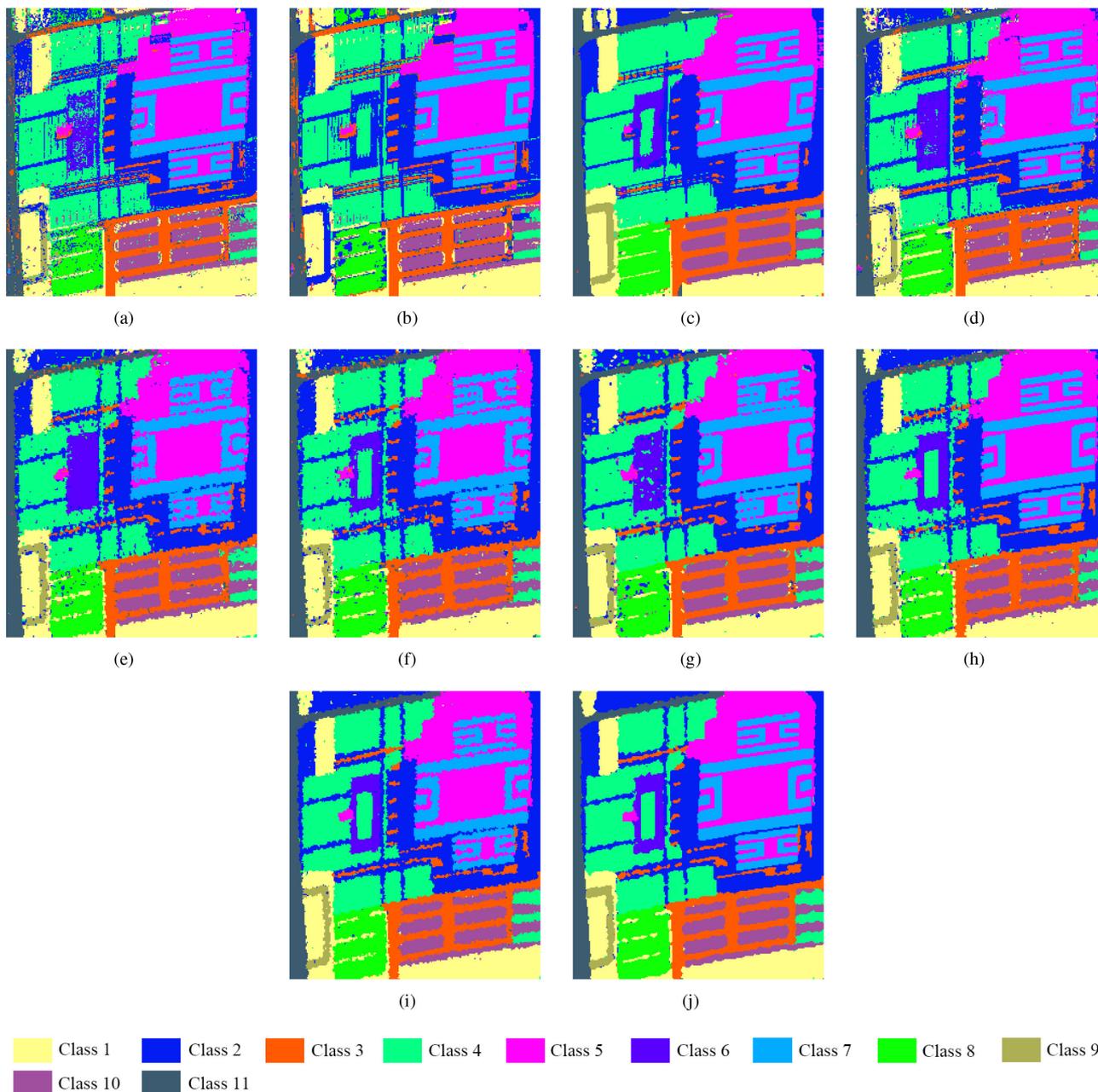
exceeds 0.95. In particular, the OA has reached 99% for PaviaU and Salinas images, and it is higher than 95% for all categories of above two images. Experimental results illustrate that almost any samples are truly classified, and the discrimination ability is enhanced to analyze the samples with similar feature values. Although the OA via deep learning model is close to that of the proposed W^2 JSRE model, the process will take more than 2000 s to complete classification for XionganS image, and it is difficult to satisfy real-time processing. As shown in Figs. 6–10, the classified noise is obviously appeared via WKNN and WSVM, which makes it difficult to recognize different objects from the images, where Grass and Vegetation categories are confused

because of the similar spectral characteristics. The JSR-based techniques are able to obtain better classification performance, and the classified noise is eliminated to some extent, but the mis-classification still exists on the edge region. The classification maps of WJSR clearly reflect different objects and correspond to the reference maps. In addition, ensemble learning is efficient to comprehensively judge the category label by a series of subclassifiers, and the objects are continuously presented for each category by using five subclassifiers. However, the learning ability is not sufficient as lack of training samples and inadequate of subclassifiers, and scattered noise is reflected on the classification maps. As for the curve of Fig. 11, the OA is

TABLE IX
OA AND KAPPA COEFFICIENT FOR INDIAN IMAGE

Class number Meas.	WKNN -	WSVM -	FCN -	DSAE -	JSR -	LAJSR -	JSR-CC -	WJSR -	W^{2}JSRE 3 sub-classifiers	W^{2}JSRE 5 sub-classifiers
1	0.0000	23.9130	50.0000	50.0000	39.1304	43.4783	43.4783	52.1739	52.1739	54.3478
2	63.5154	76.1204	71.6387	81.6527	80.7423	85.2241	85.7843	85.5042	86.4146	88.7255
3	51.0843	48.3133	77.1084	79.6386	82.6506	83.8554	85.5422	85.6627	86.2651	91.8072
4	42.6160	46.8354	98.3122	57.8059	78.9030	86.0759	53.1646	60.3376	74.6835	87.3418
5	81.9876	87.5776	92.3395	89.0269	92.1325	93.7888	97.5155	92.9607	94.4099	95.6522
6	96.9863	97.8082	96.5753	95.7534	98.6301	96.5753	97.8082	97.6712	99.1781	99.4521
7	21.4286	28.5714	100.0000	96.4286	92.8571	92.8571	96.4286	96.4286	96.4286	100.0000
8	98.7448	99.7908	100.0000	96.2343	96.8619	93.9331	98.3264	98.7448	100.0000	100.0000
9	5.0000	0.0000	100.0000	50.0000	45.0000	45.0000	50.0000	60.0000	60.0000	70.0000
10	57.9218	75.1029	81.2757	88.5802	86.6255	89.7119	91.4609	88.5802	89.9177	94.0329
11	87.8208	91.1609	76.7821	85.2546	88.7984	90.4277	90.5092	92.4644	93.0346	96.3340
12	42.8331	68.1282	80.4384	74.7049	76.2226	77.0658	78.5835	83.8111	84.6543	87.0152
13	95.6098	98.5366	99.5122	98.5366	98.5366	98.0488	98.5366	97.0732	97.5610	99.5122
14	96.9170	97.7866	95.6522	97.1542	96.4427	94.5455	97.3123	96.2846	97.4704	99.2885
15	22.7979	51.8135	90.6736	79.2746	72.2797	77.2021	79.7928	82.9016	84.7150	88.0829
16	86.0215	89.2473	100.0000	94.6237	95.6989	91.3978	95.6989	93.5484	94.6237	95.6989
OA(%)	87.3532	90.8014	92.2901	93.6361	93.9310	94.6635	94.8062	94.9156	95.1507	96.5750
Kappa	0.8201	0.8694	0.8909	0.9097	0.9140	0.9243	0.9264	0.9279	0.9308	0.9523
Time	22.3046	19.1473	857.1985	350.9767	22.8735	82.6000	108.9561	21.5868	34.5389	54.5102

TABLE X
OA AND KAPPA COEFFICIENT FOR SALINAS IMAGE

Class number Meas.	WKNN -	WSVM -	FCN -	DSAE -	JSR -	LAJSR -	JSR-CC -	WJSR -	W^{2}JSRE 3 sub-classifiers	W^{2}JSRE 5 sub-classifiers
1	98.6560	99.1538	99.0045	99.7511	99.8009	100.0000	100.0000	100.0000	100.0000	100.0000
2	99.9732	100.0000	98.0140	99.7585	99.4632	99.7585	99.8390	99.8390	99.8658	99.8658
3	97.9251	99.5445	99.2409	99.0891	99.5445	99.7470	99.2409	99.8482	99.9494	100.0000
4	99.1392	99.6413	98.9957	99.4978	99.0674	99.0674	98.8522	99.4261	99.7131	99.7131
5	98.9171	98.5437	94.4361	98.9544	99.5892	99.4772	99.8133	99.6639	99.7386	99.9253
6	99.8990	99.9242	99.7474	99.9747	99.8737	99.8484	100.0000	100.0000	100.0000	100.0000
7	99.5809	99.5529	97.6809	99.5809	99.8004	99.7206	100.0000	100.0000	100.0000	100.0000
8	87.7562	91.4559	84.2250	87.1795	86.7447	91.9617	92.0415	92.6537	95.0315	97.6400
9	99.5486	99.9355	97.9365	99.8227	99.9194	99.8872	99.9678	99.9678	99.9678	99.9839
10	94.7834	94.4173	96.9189	97.3154	97.6815	98.8408	98.2611	98.8713	99.3594	99.5424
11	97.3783	97.0037	95.8801	96.9101	99.3446	99.9064	99.8127	99.9064	99.9064	99.9064
12	100.0000	99.6886	99.3254	99.8443	100.0000	100.0000	100.0000	100.0000	100.0000	100.0000
13	97.2707	98.0349	99.1266	99.3450	99.1266	99.3450	99.1266	99.2358	99.3450	99.8908
14	94.4860	93.6449	99.8131	96.2617	96.3551	99.1589	99.5327	99.5327	99.5327	99.6262
15	71.2163	59.6175	84.4662	79.5680	84.2873	87.6170	90.1899	89.7633	92.5564	96.1062
16	98.8932	98.6165	98.2291	98.5612	99.5019	99.0592	99.8893	99.8340	99.8893	99.8893
OA(%)	96.4610	96.1091	96.7598	97.1018	97.4357	98.2530	98.4447	98.5158	98.8578	99.4744
Kappa	0.9503	0.9454	0.9546	0.9593	0.9640	0.9755	0.9782	0.9792	0.9819	0.9926
Time	304.8201	274.9726	3753.3286	1547.4970	312.6033	1114.7423	1352.6556	295.8748	473.3997	703.1399

TABLE XI
OA AND KAPPA COEFFICIENT FOR XIONGANS IMAGE

Class number Meas.	WKNN -	WSVM -	FCN -	DSAE -	JSR -	LAJSR -	JSR-CC -	WJSR -	W^{2}JSRE 3 sub-classifiers	W^{2}JSRE 5 sub-classifiers
1	90.5081	95.3314	99.4068	97.2401	95.5120	94.1450	93.3712	94.8156	95.2283	97.8334
2	93.3922	96.3916	95.7978	96.9093	98.0664	98.7058	98.5536	97.7771	98.0359	98.7820
3	71.1104	81.6417	96.5638	93.7321	90.6459	91.9504	93.2867	95.6093	96.5638	98.0910
4	93.6757	96.1382	96.4321	97.6074	96.8658	96.3061	97.9852	97.2576	97.6634	98.4749
5	97.5600	99.6950	97.5219	99.6569	99.9238	99.3138	98.7038	99.4281	99.4663	99.7331
6	96.7972	94.6619	99.0510	97.0344	96.7972	88.8493	89.3238	94.3060	94.3060	94.3060
7	77.3996	95.1605	98.0965	96.2575	89.5467	93.0795	93.9345	94.7411	96.7253	97.7900
OA(%)	87.9803	94.6579	97.1941	96.6875	95.0789	95.4934	96.0296	96.4803	96.9342	97.9013
Kappa	0.8539	0.9349	0.9659	0.9597	0.9401	0.9451	0.9516	0.9571	0.9609	0.9744
Time	63.1119	58.7080	2362.4237	945.8116	66.3752	245.1955	297.5373	61.9700	99.1521	161.5853

improved as the percentage increase of training samples, and it keeps stable on a high level as most of noise eliminated, but it is difficult to reflect a further improvement as the percentage is reached 10%, and the extent is only 0.4% as more than 10% of pixels acted as training samples. In short, the proposed W^2 JSRE model is suitable for some practical work of HSI classification, and the classification maps are well coincided with the reference maps.

TABLE XII
OA AND KAPPA COEFFICIENT FOR XIONGANU IMAGE

| Class number Meas.	WKNN -	WSVM -	FCN -	DSAE -	JSR -	LAJSR -	JSR-CC -	WJSR -	W^{2}JSRE 3 sub-classifiers	W^{2}JSRE 5 sub-classifiers
1	90.3148	90.7496	91.1394	92.4288	90.6297	89.8651	89.9700	91.0195	91.8741	93.5682
2	78.8173	77.0139	79.4123	89.6558	86.0582	87.9074	88.7221	89.0608	91.0564	93.5829
3	81.3183	79.7342	81.3365	88.5470	81.2637	80.8631	88.5834	86.4166	88.1282	92.4071
4	85.5264	89.0093	91.9947	88.0675	87.8010	92.4211	90.6353	92.9187	94.8645	97.3079
5	88.1750	87.9992	92.2822	93.2961	87.8543	87.6785	87.8854	97.5378	98.1484	98.4378
6	56.0028	1.2712	81.3616	87.3588	93.3616	87.7119	87.2175	91.7373	92.0904	93.0791
7	95.9699	95.4013	93.7291	95.4181	93.4783	93.1271	96.9565	97.4582	97.7258	99.5652
8	78.4554	79.3576	96.3190	90.5810	90.7615	91.1945	90.8336	90.4367	94.1898	96.1025
9	58.3569	0.0000	95.8924	74.2210	84.5609	80.5949	90.6516	90.3683	91.0765	95.8924
10	91.3724	94.4067	95.7619	92.8607	90.4642	90.1110	93.5923	90.8426	92.0535	92.7346
11	98.2139	71.9705	96.6807	94.3361	92.8082	91.8599	92.7028	93.7829	95.5743	96.1802
OA(%)	84.5434	82.5128	89.5003	90.9818	88.4674	89.1970	90.4741	92.6357	93.5010	95.6721
Kappa	0.8234	0.7991	0.8801	0.8971	0.8685	0.8766	0.8913	0.9159	0.9233	0.9505
Time	358.3484	322.7080	5046.2775	2079.9338	372.9808	1303.4994	1580.9001	352.5235	564.0326	839.5779
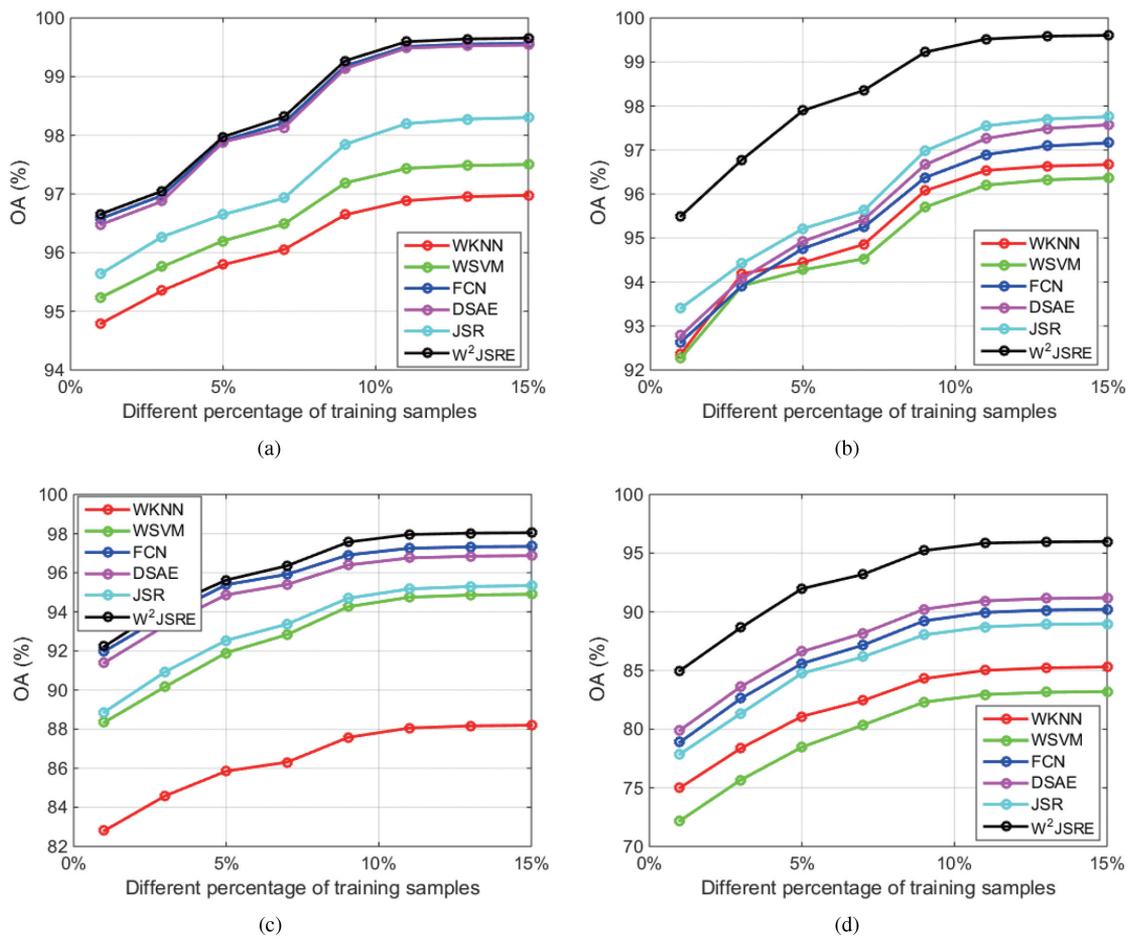
Fig. 11. OA of different percentage of training samples. (a) PaviaU image. (b) Salinas image. (c) XionganS image. (d) XionganU image.

V. CONCLUSION

In the article, an HSI classification technique based on the W^2 JSRE model and the β-WOA is proposed. The category label of each pixel is obtained by reconstructed error minimization of JSR, and the wavelet function is acted as the kernel of JSR. Moreover, ensemble learning is used to conduct detailed analysis of independent features, and the β-WOA is utilized to obtain the optimal weight of subclassifiers. In general, it is observed that the swarm intelligence algorithm is adapted to achieve the suitable weight and represent the contribution of each subclassifier. In particular, the β-WOA has the highest fitness value among the algorithms, which is appropriate to synthesize the discrimination ability of five subclassifiers. Furthermore, the optimal weight is employed to obtain the category label of HSIs, and the OA is compared with some newly proposed and corresponding HSI classification techniques. In all, the proposed W^2 JSRE model

recognizes different objects on the image, and it is sufficient to distinguish most of similar objects, which has reached 95% for pixel-level classification. As a summary, JSR combined with the wavelet kernel has a good property to solve the classification problem in most cases, the misclassification is apparently weaken by ensemble learning, and the weight optimized by the β-WOA is reasonable to improve the OA to some extent. In the future, it is preferable to combine the spatial and spectral features with different types of subclassifier for HSI classification.

REFERENCES

[1] S. Veraverbeke *et al.*, "Hyperspectral remote sensing of fire: State-of-the-art and future perspectives," *Remote Sens. Environ.*, vol. 216, pp. 105–121, 2018.

[2] M. J. Khan, H. S. Khan, A. Yousaf, K. Khurshid, and A. Abbas, "Modern trends in hyperspectral image analysis: A review," *IEEE Access*, vol. 6, pp. 14118–14129, 2018.

[3] J. M. Amigo, H. Babamoradi, and S. Elcoroaristizabal, "Hyperspectral image analysis. A tutorial," *Analytica Chimica Acta*, vol. 896, pp. 34–51, 2015.

[4] H. Liu, H. Zhu, Z. Li, and G. Yang, "Quantitative analysis and hyperspectral remote sensing of the nitrogen nutrition index in winter wheat," *Int. J. Remote Sens.*, vol. 41, pp. 858–881, 2020.

[5] Y. Wan, X. Hu, Y. Zhong, A. Ma, L. Wei, and L. Zhang, "Tailings reservoir disaster and environmental monitoring using the UAV-ground hyperspectral joint observation and processing: A case of study in Xinjiang, the belt and road," in *Proc. Int. Geosci. Remote Sens. Symp.*, 2019, pp. 9713–9716.

[6] M. L. Clark and N. E. Kilham, "Mapping of land cover in northern California with simulated hyperspectral satellite imagery," *ISPRS J. Photogrammetry Remote Sens.*, vol. 119, pp. 228–245, 2016.

[7] L. He, J. Li, C. Liu, and S. Li, "Recent advances on spectral-spatial hyperspectral image classification: An overview and new guidelines," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 3, pp. 1579–1597, Mar. 2018.

[8] Y. Yuan, X. Zheng, and X. Lu, "Hyperspectral image superresolution by transfer learning," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 10, no. 5, pp. 1963–1974, May 2017.

[9] E. M. Tzanakou, *Supervised and Unsupervised Pattern Recognition: Feature Extraction and Computational Intelligence*. Boca Raton, FL, USA: CRC Press, 2017.

[10] M. Zhang, J. Ma, and M. Gong, "Unsupervised hyperspectral band selection by fuzzy clustering with particle swarm optimization," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 5, pp. 773–777, May 2017.

[11] B. Barman and S. Patra, "Variable precision rough set based unsupervised band selection technique for hyperspectral image classification," *Knowl.-Based Syst.*, vol. 193, 2020, Art. no. 105414.

[12] P. Hong, "An improved hyperspectral image classification approach based on ISODATA and SKR method. infrared, millimeter-wave, and terahertz technologies IV," Int. Soc. Opt. Photon., vol. 10030, 2016, Art. no. 100301W

[13] S. Patra, K. Bhardwaj, and L. Bruzzone, "A spectral-spatial multicriteria active learning technique for hyperspectral image classification," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 10, no. 12, pp. 5213–5227, Dec. 2017.

[14] J. Xia, N. Falco, J. A. Benediktsson, P. Du, and J. Chanussot, "Hyperspectral image classification with rotation random forest via KPCA," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 4, pp. 1601–1609, Apr. 2017.

[15] W. Sun, C. Liu, Y. Xu, L. Tian, and W. Li, "A band-weighted support vector machine method for hyperspectral imagery classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 10, pp. 1710–1714, Oct. 2017.

[16] J. Liu, Z. Wu, J. Li, L. Xiao, A. Plaza, and J. A. Benediktsson, "Spatial-spectral hyperspectral image classification using random multiscale representation," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 9, no. 9, pp. 4129–4141, Sep. 2016.

[17] H. Huang, Y. Duan, H. He, G. Shi, and F. Luo, "Spatial-spectral local discriminant projection for dimensionality reduction of hyperspectral image," *ISPRS J. Photogrammetry Remote Sens.*, vol. 156, pp. 77–93, 2019.

[18] M. E. Paoletti, J. M. Haut, J. Plaza, and A. Plaza, "Deep learning classifiers for hyperspectral imaging: A review," *ISPRS J. Photogrammetry Remote Sens.*, vol. 158, pp. 279–317, 2019.

[19] Z. Zhang, Y. Xu, J. Yang, X. Li, and D. Zhang, "A survey of sparse representation: Algorithms and applications," *IEEE Access*, vol. 3, pp. 490–530, 2015.

[20] S. Chen, C. Guo, and J. Lai,, "Deep ranking for person re-identification via joint representation learning," *IEEE Trans. Image Process.*, vol. 25, no. 5, pp. 2353–2367, May 2016.

[21] K. Tan, J. Hu, J. Li, and P. Du, "A novel semi-supervised hyperspectral image classification approach based on spatial neighborhood information and classifier combination," *ISPRS J. Photogrammetry Remote Sens.*, vol. 105, pp. 19–29, 2015.

[22] J. Peng and Q. Du, "Robust joint sparse representation based on maximum correntropy criterion for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 12, pp. 7152–7164, Dec. 2017.

[23] J. Peng, X. Jiang, N. Chen, and H. Fu, "Local adaptive joint sparse representation for hyperspectral image classification," *Neurocomputing*, vol. 334, pp. 239–248, 2019.

[24] B. Tu, X. Zhang, X. Kang, G. Zhang, J. Wang, and J. Wu, "Hyperspectral image classification via fusing correlation coefficient and joint sparse representation," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 3, pp. 340–344, Mar. 2018.

[25] E. Zhang, X. Zhang, L. Jiao, H. Liu, S. Wang, and B. Hou, "Weighted multifeature hyperspectral image classification via kernel joint sparse representation," *Neurocomputing*, vol. 178, pp. 71–86, 2016.

[26] X. Zhou, J. Sun, X. Wu, B. Lu, N. Yang, C. Dai, "Research on moldy tea feature classification based on WKNN algorithm and NIR hyperspectral imaging," *Spectrochimica Acta Part A: Mol. Biomol. Spectrosc.*, vol. 206, pp. 378–383, 2019.

[27] M. Wang, C. Wu, L. Wang, D. Xiang, and X. Huang, "A feature selection approach for hyperspectral image based on modified ant lion optimizer," *Knowl.-Based Syst.*, vol. 168, pp. 39–48, 2019.

[28] J. Xia, P. Ghamisi, N. Yokoya, and A. Iwasaki, "Random forest ensembles and extended multiextinction profiles for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 1, pp. 202–216, Jan. 2018.

[29] B. Pan, Z. Shi, and X. Xu, "Hierarchical guidance filtering-based ensemble classification for hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 7, pp. 4177–4189, Jul. 2017.

[30] L. Liu, L. Chen, C. L. P. Chen, Y. Y. Tang, and C. M. pun, "Weighted joint sparse representation for removing mixed noise in image," *IEEE Trans. Cybern.*, vol. 47, no. 3, pp. 600–611, Mar. 2017.

[31] M. Taherkhani and R. Safabakhsh, "A novel stability-based adaptive inertia weight for particle swarm optimization," *Appl. Soft Comput.*, vol. 38, pp. 281–295, 2016.

[32] S. Mirjalili and A. Lewis,, "The whale optimization algorithm," *Adv. Eng. Softw.*, vol. 95, pp. 51–67, 2016.

[33] I. Aljarah, H. Faris, and S. Mirjalili,, "Optimizing connection weights in neural networks using the whale optimization algorithm," *Soft Comput.*, vol. 22, pp. 1–15, 2018.

[34] X. Xu, H. Rong, M. Trovati, M. Liptrott, and N. Bessis, "CS-PSO: Chaotic particle swarm optimization algorithm for solving combinatorial optimization problems," *Soft Comput.*, vol. 22, pp. 783–795, 2018.

[35] V. S. Ozsoy, M. G. Unsal, and H. H. Orkcu, "Use of the heuristic optimization in the parameter estimation of generalized gamma distribution: Comparison of GA, DE, PSO and SA methods," *Comput. Statist.*, vol. 374, pp. 1–31, 2020.

[36] Y. Yuan, J. Lin, and Q. Wang,, "Hyperspectral image classification via multitask joint sparse representation and stepwise MRF optimization," *IEEE Trans. Cybern.*, vol. 46, no. 12, pp. 2966–2977, Dec. 2016.

[37] M. M. Mafarja and S. Mirjalili,, "Hybrid whale optimization algorithm with simulated annealing for feature selection," *Neurocomputing*, vol. 260, pp. 302–312, 2017.

[38] L. Li, P. Liu, Y. Xing, and H. Guo, "Time-frequency analysis of acoustic signals from a high-lift configuration with two wavelet functions," *Appl. Acoust.*, vol. 129, pp. 155–160, 2018.

[39] X. Li, Z. Xu, W. Cai, and X. Shao, "Filter design for molecular factor computing using wavelet functions," *Analytica Chimica Acta*, vol. 880, pp. 26–31, 2015.

[40] L. Dong and J. Liao,, "Wavelet kernel function based multiscale LSSVM for elliptic boundary value problems," *Neurocomputing*, vol. 356, pp. 40–51, 2019.

[41] Y. Blache, C. Hautier, F. Lefebvre, A. Djordjevic, T. Creveaux, and I. Rogowski, "Analysis of the tennis racket vibrations during forehand drives: Selection of the mother wavelet," *J. Biomech.*, vol. 61, pp. 94–101, 2017.

[42] M. Grana, M. Veganzons, and B. Ayerdi, Resource of HSI. Accessed: Jan. 8, 2021. [Online]. Available: http://www.ehu.eus/ccwintco/index.php

[43] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Proc. Int. Conf. Neural Netw.*, 1995, vol. 4, pp. 1942–1948.
[44] K. V. Price, "Differential evolution: A fast and simple numerical optimizer," *Proc. North Amer. Fuzzy Inf. Process.*, 1996, pp. 524–527.
[45] X. S. Yang and S. Deb, "Cuckoo search via Levy flights," in *Proc. World Congr. Nat. Biol. Inspired Comput.*, 2009, pp. 210–214.
[46] S. Mirjalili, S. M. Mirjalili, and A. Lewis, "Grey wolf optimizer," *Adv. Eng. Softw.*, vol. 69, pp. 46–61, 2014.
[47] S. Mirjalili, "The ant lion optimizer," *Adv. Eng. Softw.*, vol. 83, pp. 80–98, 2015.
[48] C. Witt, "Tight bounds on the optimization time of a randomized search heuristic on linear functions," *Combinatorics, Probab. Comput.*, vol. 22, pp. 294–318, 2013.
[49] L. Zou, X. Zhu, C. Wu, Y. Liu, and L. Qu, "Spectral–spatial exploration for hyperspectral image classification via the fusion of fully convolutional networks," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 659–674, 2020.
[50] P. Zhou, J. Han, G. Cheng, and B. Zhang, "Learning compact and discriminative stacked autoencoder for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 7, pp. 4823–4833, Jul. 2019.

Maolin Chen received the B.S., M.S., and Ph.D. degrees from Wuhan University, Wuhan, China, in 2012, 2014, and 2018, respectively, all in photogrammetry and remote sensing.

He is currently an Associate Professor with the School of Civil Engineering, Chongqing Jiaotong University, Chongqing, China. His research interests include feature extraction, image interpretation, and object recognition of laser scanning data.

Shuping Wang received the B.S. and M.S. degrees from the Hubei University of Technology, Wuhan, China, in 2011 and 2014, respectively, both in computer science and technology.

Since 2015, he has been an Engineer with the Hubei Cancer Hospital, Huazhong University of Science and Technology, Wuhan. His major research interests include medical image processing, swarm intelligence algorithm, and computer application technology.

Mingwei Wang received the B.S. degree in electronic information science and technology from Hubei Normal University, Huangshi, China, in 2011, the M.S. degree in software engineering from the Hubei University of Technology, Wuhan, China, in 2015, and the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, in 2018.

Since 2018, he has been a Professional Researcher with the Institute of Geological Survey, China University of Geosciences, Wuhan. His major research interests include hyperspectral image processing, swarm intelligence algorithm, and deep learning models.

Zhiwei Ye received the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 2006.

He is currently a Professor with the School of Computer Science, Hubei University of Technology, Wuhan. He has authored or coauthored more than 30 papers in the area of digital image processing and swarm intelligence algorithm. His research interests include image analysis, pattern recognition, and data mining.

Zitong Jia received the B.S. degree in hydrology and water resources engineering from Northwest Agriculture and Forest University, Xianyang, China, in 2020. She is currently working toward the M.S. degree with the China University of Geosciences, Wuhan, China.

Her major research interests include hydrologic model research based on geographic information system, remote sensing image processing, and classification of land-cover change.

Jianwei Luo received the B.S. degree in computer science and technology from the Wuhan University of Technology, Wuhan, China, in 2005.

Since 2012, he has been a Senior Engineer with the Hubei Cancer Hospital, Huazhong University of Science and Technology, Wuhan. His major research interests include hyperspectral imaging theory, machine learning models, and computer application technology.