Applied Intelligence (2021) 51:7766–7780
https://doi.org/10.1007/s10489-021-02270-0

A band selection approach based on wavelet support vector machine ensemble model and membrane whale optimization algorithm for hyperspectral image

Mingwei Wang[1,2] · Ziqi Yan[1] · Jianwei Luo[3] · Zhiwei Ye[4] · Peipei He[5]

Accepted: 10 February 2021 / Published online: 17 March 2021

Abstract

Hyperspectral Image (HSI) has become one of the important remote sensing sources for object interpretation by its abundant band information. Among them, band selection is considered as the main theme in HSI classification to reduce the data dimension, and it is a combinatorial optimization problem and difficult to be completely solved by previous techniques. Whale Optimization Algorithm (WOA) is a newly proposed swarm intelligence algorithm that imitates the predatory strategy of humpback whales, and membrane computing is able to decompose the band information into a series of elementary membranes that decreases the coding length. In addition, Support Vector Machine (SVM) combined with wavelet kernel is adapted to HSI datasets with high dimension and small samples, ensemble learning is an effective tool that synthesizes multiple sub-classifiers to solve the same problem and obtains accurate category label for each sample. In the paper, a band selection approach based on wavelet SVM (WSVM) ensemble model and membrane WOA (MWOA) is proposed, experimental results indicate that the proposed HSI classification technique is superior to other corresponding and newly proposed methods, achieves the optimal band subset with a fast convergence speed, and the overall classification accuracy has reached 93% for HSIs.

Keywords Hyperspectral image · Band selection · Whale optimization algorithm · Membrane computing · Classifier ensemble · Wavelet support vector machine

1 Introduction

As a sensor type with enough features, hyperspectral remote sensing has been applied on a series of aspects with its dense sampling of narrow and continuous band information, each band expresses a one-dimensional feature for the dataset, and the abundant spectral resolution supplies the potential for precisely discrimination of different objects [1, 2]. Hyperspectral image (HSI) classification is projected to obtain the category label of each pixel on HSIs and conduct object analysis for the whole image, such as building, vegetation, and water [3]. In general, they are separated into two scopes: unsupervised and supervised techniques. For unsupervised techniques, fuzzy clustering model, graph-based clustering, and iterative self-organizing data analysis techniques algorithm (ISODATA) have been utilized to obtain the object information of HSIs [4–6]. However, there is no prior knowledge for the above approaches, which blindly conduct classification with the characteristics of feature values itself, and satisfactory classification results are obtained only under the special conditions [7]. For supervised techniques, active learning, Bayesian modelling, neural network have been utilized to conduct classification for HSIs [8–10]. In short, the supervised techniques need to know the characteristics of training samples in advance, and higher classification accuracy is achieved than unsupervised techniques in most situations.

As a frequently used supervised classification technique, support vector machine (SVM) is an active machine learning model to solve the classification problem on the dataset with high dimension and small samples. It is on the basis of experiential learning theory and structural risk minimization principle and has been extensively applied in a great deal of fields [11, 12]. In particular, several of studies have explored the problem of HSI classification by using SVM, and preferable classification accuracy is achieved

✉ Jianwei Luo
wuljw@126.com

Extended author information available on the last page of the article.

than traditional supervised classification techniques [13, 14]. However, as the improvement of spatial resolution, the objects become more specified in the local region, and it is difficult to discriminate different objects with similar feature values by using single classifier [15]. Ensemble learning is a paradigm of machine learning that synthesizes multiple sub-classifiers to solve the same problem, and better generalization ability is obtained than single classifier according to different emphasis of sub-classifiers especially for indeterminate objects [16]. As a result, classifier ensemble based on SVM has been utilized to solve financial distress and druggable proteins problems with dozens of features on public datasets [17, 18]. As for HSI classification, the computational efficiency is limited as all of bands directly input into each sub-classifier [19–21]. Moreover, kernel function is the core issue for the performance of SVM, which is fused to map the input into solving space with higher dimension, so that the complex classification problem is transferred to linearly separable [22]. In recent years, the uniting of wavelet theory and SVM, namely, wavelet SVM (WSVM) has been built as an improvement of SVM, the feature vectors are extracted from time series because of positive non-linear mapping of SVM and locally analyze of wavelet kernel, but the mapping into trigonometric function with a fixed type may not adapt to different distributions of datasets [23].

On the other hand, the abundant band information of HSIs also leads to the curse of dimensionality, and band selection is projected to select a series of feature values corresponding to band information from HSIs that composes a band subset to avoid data redundancy and obtain higher classification accuracy [24, 25]. The generation of band subset is a search behavior that chooses the combination of items from HSIs via complete, random or heuristic search strategy. If a HSI contains N bands, $O(2^N)$ possible subsets will be generated, which is considered as a non-polynomial hard problem. As a combinatorial optimization problem, swarm intelligence algorithms with heuristic search have been utilized as guiding factors to seek for the solution of band selection [26]. For instance, Su et al. [27] and Ghosh et al. [28] proposed a series of band selection techniques integrating SVM with particle swarm optimization (PSO) and differential evolution (DE) to achieve the optimal band subset via satisfactory classification accuracy, but the optimization ability was limited due to the lack of local search. In recent years, some newly proposed swarm intelligence algorithms had been proposed and applied on the aspect of band selection, such as gravitational search algorithm (GSA), Gray Wolf Optimizer (GWO), ant lion optimizer (ALO), et al. [29–31], which were capable of obtaining superior outcomes in tackling band selection problem when compared with the previous techniques, but the time complexity was increased as the improvement of spectral resolution and total number of bands.

Whale Optimization Algorithm (WOA) is a newly proposed heuristic search algorithm and has been widely used in diverse applications [32–34]. The performance of WOA depends on only one parameter, which makes it not easy to trap into the local optima, and it has stable performance that converges to the global optimal solution. However, the coding length of each individual is equal to the number of bands for band selection, and CPU time is uncontrolled to search for the global optimal solution from a large number of candidate solutions when all of bands input into each sub-classifier. Membrane computing is a novel paradigm of natural computing, the long coding can be separated into a series of short coding that enter into elementary membranes [35], and an elementary membrane is corresponding to a sub-classifier. Where the optimal band combination of each sub-classifier is obtained by WOA, then transmit to the skin, and the optimal band subset of HSI datasets is achieved. Hence, a band selection approach based on WSVM ensemble model and membrane WOA (MWOA) is proposed to obtain the category label of each pixel, the HSI datasets are separated into several parts by no duplicate selection of band information, and inputs into corresponding sub-classifiers. Moreover, the ensemble with WSVM for each sub-classifier concerns on the homogeneity and the ensemble with multiple wavelet kernels emphasize the heterogeneity, the original dataset is mapped into quadratic, exponential, trigonometric functions with different types, which is suitable for different distributions of datasets especially for HSIs. The voting strategy is replaced by the information integration, and the category label of each sample is output by the remaining band information on the skin of MWOA. The main contributions of the paper are displayed as follows:

- A novel WSVM ensemble model is designed, and WSVM with different wavelet kernels is utilized to balance the heterogeneity and homogeneity at the same time.
- The optimal band combination of each sub-classifier is obtained by WOA, and a novel MWOA is presented to obtain the optimal band subset of HSI datasets according to the coding on the skin.
- The category label of each sample is output by the remaining band information on the skin, and the discrimination ability of different objects with similar feature values is enhanced.

The rest of paper is structured as below. Section 2, describes the backgrounds of WSVM and WOA. The process of proposed technique for band selection is exhibited in Section 3. Section 4, sums up the experimental results and expends data discussion. Finally, the paper is concluded in Section 5.

2 Backgrounds

2.1 Basic theory of SVM

Assume that a classification problem consists of n instance-label pairs, $S = (x_i, y_i), (i = 1, 2, ..., n)$, $x_i \in R^a$ is an instance vector and $y_i \in \{-1, +1\}$ is the category label of samples. The process of training is to search for a hyper-plane that separates the positive (+1) samples from the negative (-1) samples, which is resolved by the optimization of following expression:

$$\phi(\omega, \varepsilon) = \frac{1}{2}||\omega||^2 + C\sum_{i=1}^{n}\varepsilon_i \tag{1}$$

Subject to constraints:

$$y_i[(\omega \cdot x_i) + b] \geq 1 - \varepsilon_i, i = 1, 2, ..., n \tag{2}$$

where ω is the normal vector of hyper-plane, $\varepsilon_i \geq 0$ is the slack variable to enhance the samples that are fault classified, C is the penalty factor to balance the error term $\sum_{i=1}^{n}\varepsilon_i$, and ϕ is the kernel function that maps the input into another solving space with higher dimension, and the transformation of space depends on the definition of kernel, which is expressed as (3):

$$K(x, x') = \phi(x)^T \cdot \phi(x') \tag{3}$$

Among them, the widely adopted kernel is radial basis function (RBF) that is defined as (4):

$$K(x, x') = exp\left[\frac{-(x_i - x_i')^2}{2\sigma_i^2}\right] \tag{4}$$

However, the transformation of space is conducted on a fixed scale that may not discriminate different forms of input especially for similar characterization, and it is insufficient to generate a linearly-separable space about similar feature values and high dimensional vectors.

2.2 Wavelet kernels combined with SVM

A kernel function that obeys Mercer's theorem is known as an admissible Support Vector (SV), it is decomposed as shift-invariant form $K(x, x') = K(x - x')$ in the solving space, and Mercer's theorem offers essential factors to determine whether the shift-invariant form is an admissible SV kernel [36].

The basic theory of wavelet analysis is to combine a lot of linear wavelet bases that reflect an arbitrary function $f(x)$. Assume that $\phi(x)$ is a function of one-dimensional mother wavelet basis, and a separable multi-dimensional wavelet function should be expressed as:

$$\phi_d(x) = \prod_{i=1}^{d} f(x_i) \tag{5}$$

The shift-invariant form can be constructed for wavelet function as follows:

$$\phi_d(x, x') = \prod_{i=1}^{d} f\left(\frac{x_i - x_i'}{\sigma_i}\right), \tag{6}$$

where $\sigma_i > 0$ is the wavelet scale factor.

Where the existing mother wavelet basis contents the fixed condition of shift-invariant form, and the following equations respectively name Mexican hat, Complex Morlet, Shannon, and Harmonic functions are considered as an admissible SV kernel, and they can be expressed as the following equations [37–40]:

$$K(x, x') = \prod_{i=1}^{d}\left[1 - \frac{(x_i - x_i')^2}{\sigma_i^2}\right] \cdot exp\left[-\frac{(x_i - x_i')^2}{2\sigma_i^2}\right], \tag{7}$$

$$K(x, x') = \prod_{i=1}^{d} cos\left[1.75 \times \frac{(x_i - x_i')}{\sigma_i}\right] \cdot exp\left[-\frac{(x_i - x_i')^2}{2\sigma_i^2}\right], \tag{8}$$

$$K(x, x') = \prod_{i=1}^{d} \frac{sin\left[\frac{\pi}{2} \cdot \frac{(x_i - x_i')}{\sigma_i}\right]}{\frac{\pi}{2} \cdot \frac{(x_i - x_i')}{\sigma_i}} \cdot cos\left[\frac{3\pi}{2} \cdot \frac{(x_i - x_i')}{\sigma_i}\right]. \tag{9}$$

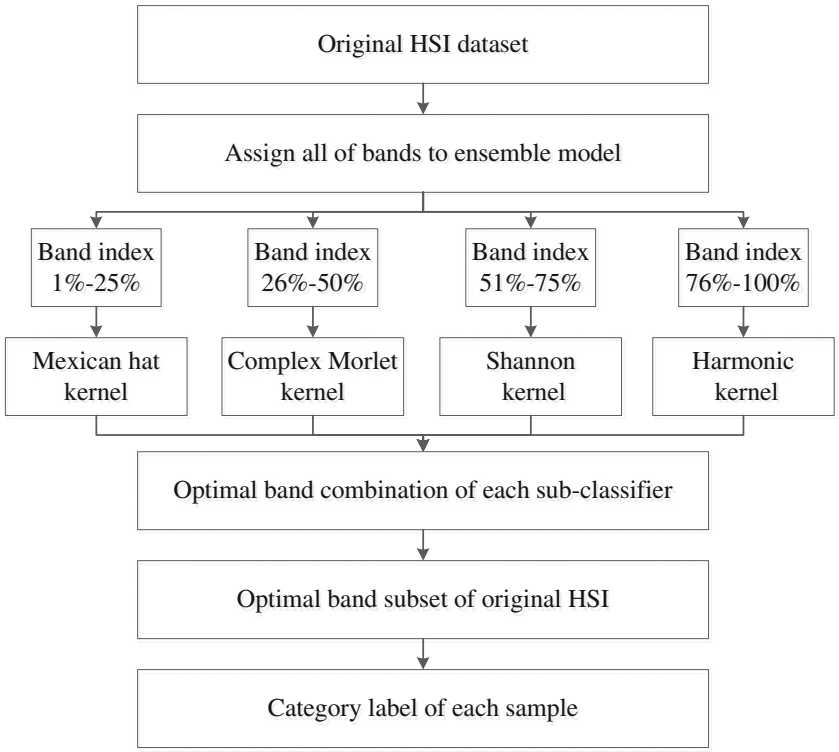

Fig. 1 WSVM ensemble model for band selection

$$K(x, x') = \prod_{i=1}^{d} \frac{e^{i4\pi \frac{x_i - x_i'}{\sigma_i}} - e^{i2\pi \frac{x_i - x_i'}{\sigma_i}}}{i2\pi \left(\frac{x_i - x_i'}{\sigma_i}\right)} \tag{10}$$

The above wavelet functions not only possess translation orthogonality, but also approximate an arbitrary equation in the square integral space, and the transformation on multiple scales guaranteeing the input is more likely to be discriminated. Due to the wavelet function has the ability of non-linear mapping on different scales, WSVM is adapted for classification decision-making and pays attention on misclassification samples.

2.3 Mathematical model of WOA

Whale optimization algorithm (WOA) is based on the predatory strategy of humpback whales that tend to catch crowd of krill or small fishes near to the surface, the process

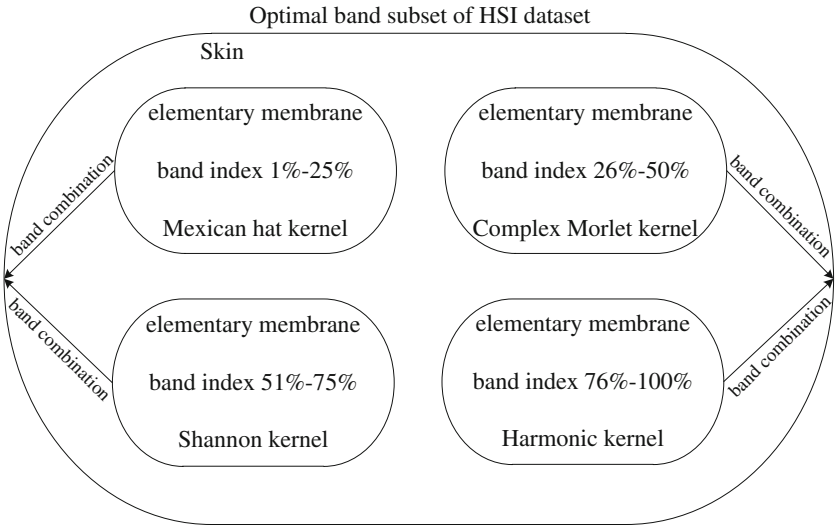

Fig. 2 Membrane computing for WSVM ensemble model

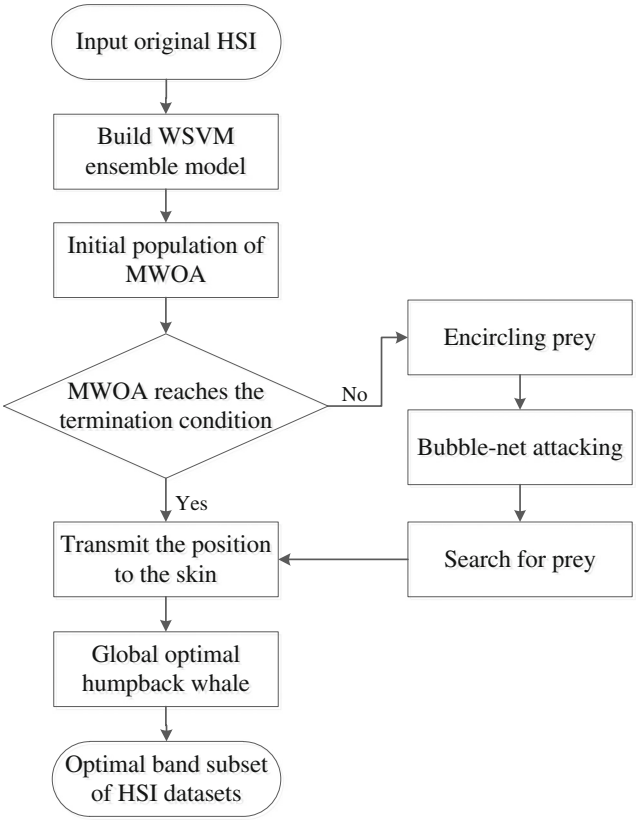

Fig. 3 Flow chart of the proposed method

is conducted by producing specific bubbles with a ring path, and the operator is separated into three parts that are encircling prey, spiral bubble-net attacking, search for prey, and the main procedure of WOA is depicted as below:

Encircling prey Humpback whales have the ability to search for the location of prey and encircle them. It is assumed that the position of current optimal solution is the target prey or it is the proximate solution to the optimum in theory. Other humpback whales should endeavour to motivate their positions towards to it. The process is written as follows:

$$\mathbf{D} = |\mathbf{C} \cdot X^*(t) - X(t)|, \tag{11}$$

$$X(t+1) = X^*(t) - \mathbf{A} \cdot \mathbf{D}, \tag{12}$$

Where t is the number of current iteration, $X^*(t)$ is the position of prey, $X(t)$ and $X(t+1)$ respectively represent the position of humpback whales in current and later procedure. $\mathbf{A}$ and $\mathbf{C}$ are the variable vectors that are expressed as $\mathbf{A} = 2\mathbf{a} \cdot \mathbf{r} - \mathbf{a}$ and $\mathbf{C} = 2 \cdot \mathbf{r}$, and $\mathbf{a}$ is decreased gradually within the scope of [2,0] and $\mathbf{r}$ is a random number with uniform distribution.

Bubble-net attacking Each humpback whale moves close to the prey within a compact ring acting the exploitation phase and follows with a spiral-shaped path in the meantime, and it is supposed that a probability of 0.5 is set to choose whether the compact ring or spiral mechanism and renew the position of humpback whale according to the distance between current humpback whale and prey. The formulation about current process is expressed as below:

$$X(t+1) = \mathbf{D}' \cdot e^{bl} \cdot cos(2\pi l) + X^*(t) \tag{13}$$

where $\mathbf{D}'$ is the distance of current humpback whale to prey, which is expressed as $\mathbf{D}' = |X^*(t) - X(t)|$, b is a constant that indicates the situation of logarithmic spiral, l is a random number within the scope of [-1,1].

Search for prey The current behavior combined with vector $\mathbf{A}$ is utilized to renew the position of prey, and the random value greater than 1 or less than -1 is set for $\mathbf{A}$ that lets the humpback whale jump out of the local space, and the position of current humpback whale is updated according to the random walk strategy rather than the best humpback whale. The details are expressed as (15):

$$\mathbf{D} = |\mathbf{C} \cdot X_{rand} - X(t)| \tag{14}$$

$$X(t+1) = X_{rand} - \mathbf{A} \cdot \mathbf{D} \tag{15}$$

where $\mathbf{D}$ is the distance between a random and current humpback whale, X_{rand} is the position of a random humpback whale selected from the whole population.

Table 1 Parameters setting for different algorithms

Parameters	Value
Population size	40
Number of runs for each algorithm	50
G_0 Initial gravitational variable in GSA	100
a Correlation coefficient in GWO	[2,0]
ω Exploration level in ALO	[2,6]
a Correlation coefficient in WOA	[2,0]

Table 2 Fitness value and classification accuracy of different algorithms

Dataset	Meas.	GSA	GWO	ALO	WOA	MGSA	MGWO	MALO	MWOA
SalinasA	Fiv	0.9251	0.9293	0.9349	0.9377	0.9425	0.9466	0.9522	0.9559
	Acc(%)	96.0477	96.3285	96.7336	97.0093	97.7785	97.9832	98.3308	98.5047
HSI1	Fiv	0.9016	0.9044	0.9079	0.9111	0.9264	0.9293	0.9329	0.9380
	Acc(%)	93.7992	93.9817	94.2209	94.4545	95.6292	95.8375	96.0292	96.2250
HSI2	Fiv	0.8512	0.8555	0.8602	0.8657	0.8890	0.8927	0.8940	0.8972
	Acc(%)	86.1939	86.5885	87.0080	87.6998	90.5194	91.1597	91.3035	92.0018
HSI3	Fiv	0.8884	0.8917	0.8958	0.8993	0.9102	0.9143	0.9181	0.9225
	Acc(%)	90.5068	90.8881	91.2002	91.6078	93.0451	93.3010	93.6625	94.1556

3 The proposed method

3.1 Structure of WSVM ensemble model

As for ensemble learning, multiple sub-classifiers are simultaneously trained and then polymerized to construct an ensemble model. A WSVM ensemble model is proposed to synthesize multiple sub-classifiers into a robust one that improves the generalization and discrimination abilities, the peculiarity of ensemble is represented by different band combination that inputs to each sub-classifier, and the construction is shown on Fig. 1.

The key issue of classifier ensemble depends on two elements that are how to construct each sub-classifier and how to fuse the sub-classifiers and build an ensemble classifier. This is conducted as follows. First, four sub-classifiers are built by WSVM and respectively corresponding to Mexican hat, Complex Morlet, Shannon and Harmonic kernels in Section 2.2. Then, all of bands are entered into WSVM ensemble model with the indexed sequential of HSIs, that is, the band indexes located at 1%-25% are assigned to first sub-classifier, the band indexes located at 26%-50% are assigned to second sub-classifier, the band indexes located at 51%-75% are assigned to third sub-classifier, and the band indexes located at latter 25% are assigned to fourth sub-classifier. Further, the optimal band combination of each sub-classifier is obtained, and the category label of each sample is output as the optimal band subset is achieved by information integration. The voting strategy is not necessary for the process of ensemble to avoid the category ambiguity of some samples with similar votes.

3.2 Strategy of MWOA

Membrane computing is a component of natural computing that investigates computing models abstracted from the interactions of several cells in tissues, and a complex problem can be decomposed to the combination of some easy solved problems. The core topic for MWOA and WSVM ensemble model is the expression form of band selection to be handled, and a suitable mapping between MWOA and WSVM ensemble model corresponding to the solution space is necessary, and the structure of membrane computing combined with WSVM ensemble model is demonstrated on Fig. 2.

Where each band has exactly two candidate statuses for the process of band selection, they are selected or deselected. If the number of bands is N for a HSI, the coding length is equal to $[N/4]$ for a humpback whale. Every bit of MWOA is set by "0" or "1", where "1" means the current band will be selected, and "0" means the current band will be unselected. For instance, there are 10 bands for a HSI and 2 sub-classifiers are used

Table 3 Selected number of bands and CPU time of different algorithms

Dataset	Meas.	GSA	GWO	ALO	WOA	MGSA	MGWO	MALO	MWOA
SalinasA	Ft	22.4000	21.9667	21.4667	21.1000	8.7667	8.2333	7.8000	7.0333
	Time	4.2616	4.3760	4.1515	3.9409	1.3889	1.4477	1.3228	1.2673
HSI1	Ft	21.5000	21.0667	20.5000	19.8667	7.1667	6.7667	6.4000	5.8333
	Time	8.2922	8.5886	8.0069	7.6565	2.8472	2.9339	2.6961	2.5050
HSI2	Ft	26.0000	25.4667	24.7667	23.9333	10.5000	10.0333	9.7000	8.9333
	Time	8.5175	8.7665	8.2999	7.8773	2.9466	3.0508	2.8428	2.6195
HSI3	Ft	23.2000	22.7667	22.0333	21.7000	9.3000	8.8333	8.3333	7.5000
	Time	9.6488	9.9567	9.2288	8.7660	3.6222	3.7588	3.5393	3.2850

for classification, and the coding on the skin for MWOA is "0010000101 (elementary membrane 1) ‖0100101000 (elementary membrane 2)". That is, the 3rd, 8th, 10th bands of the previous sub-classifier and 12th, 15th, 17th bands of latter sub-classifier will be selected for classification, and remaining bands will be abandoned. The optimal band subset of HSI datasets is obtained as the optimal individuals on elementary membranes are transmitted to the skin.

3.3 Definition of objective function

The main goal of band selection for each sub-classifier is to improve the classification accuracy and reduce the selected number of bands of each individual, and the performance of classifier ensemble is enhanced by the combination of a series of sub-classifiers. In the paper, WSVM is acted as the sub-classifier for the process of classification here. Moreover, the classification accuracy is just an important goal, and the reduction of independent bands is also an imperative goal. The comprehensive goal is to gain the higher classification accuracy combined with less number of bands as possible. Thus, the fitness value is computed as (16):

$$F(i) = 0.25 * \sum_{j=1}^{4} \left[\lambda \cdot Acc(i, j) + (1 - \lambda) \cdot log_{10} \frac{n_c}{n_s(i, j)} \right] \tag{16}$$

where $F(i)$ indicates the fitness value of i-th humpback whale, n_c and $n_s(i, j)$ are respectively the total and selected number of bands about j-th elementary membrane, and $Acc(i, j)$ is the classification accuracy of j-th sub-classifier. λ is a weighting parameter to balance the classification accuracy and selected number of bands, which is set as $\lambda = 0.9$ here.

3.4 Implementation of the proposed method

The proposed band selection approach is easy to be fulfilled, WSVM ensemble model and MWOA are designed to obtain the optimal band subset and conduct pixel-level classification for entire HSIs, and the exact process is listed as the following flow chart (Fig. 3) and pseudocode:

Algorithm 1 Band selection via WSVM ensemble model and MWOA.

Require: Construct training and testing samples based on the band information of HSIs, and the iteration number of MWOA is t = 0.

Ensure: The category label of each sample on HSIs.

1: Input the actual HSI, the feature values of some labelled samples are extracted by ENVI software;
2: Build WSVM ensemble model, assign all of bands to the corresponding sub-classifier;
3: Generate initial population of MWOA, and the coding of an elementary membrane is corresponding to a sub-classifier;
4: **while** The algorithm does not reach the termination condition **do**
5: t = t + 1;
6: Conduct testing for WSVM ensemble model, and compute the fitness value by (16);
7: Encircling prey by using (12);
8: Bubble-net attacking by using (13);
9: Search for prey by using (15);
10: Obtain the optimal humpback whale on the elementary membrane;
11: **if** The optimal humpback whale on the elementary membrane is better than that of last iteration **then**
12: The position of humpback whale on the elementary membrane is supplanted at the current state;
13: **end if**
14: Transmit the position of humpback whale to the skin, and output the position of global optimal humpback whale;
15: **end while**
16: **return** The optimal band subset of HSI datasets and the category label of each sample.

4 Experimental results and discussion

The proposed band selection approach is accomplished by the language of MATLAB 2014b on a personal computer

Table 4 OA of different HSI classification methods (%)

Image	LS-SVM	WSVM	TCRC	RPN	DSVM	WOA	MWOA
SalinasA	96.4976	97.6884	98.7812	98.9913	99.1174	99.2855	99.4396
HSI1	90.1047	92.7789	94.6462	94.4211	96.1538	97.2216	97.9784
HSI2	67.3196	71.4775	76.3307	81.3756	83.8544	92.4694	93.2659
HSI3	61.6156	69.0980	74.3854	84.3135	88.0635	94.8627	95.3538

with a 3.60 GHz CPU, 8.00G RAM under Windows 10 operation system.

The process mainly concerns on the optimization ability of MWOA and the classification performance of WSVM ensemble model, a public collected and 3 measured airborne HSIs respectively named SalinasA [41], HSI1, HSI2 and HSI3 are utilized here. The total number of bands is respectively 204 and 100 for public and measured HSIs, there are respectively 51 and 25 bands that are input into each sub-classifier, and the coding length of MWOA is also 51 or 25 for each individual. For HSI datasets extracted by ENVI software, we randomly choose 10% samples of each category as training data, and remaining 90% are selected as testing data.

Moreover, some newly proposed swarm intelligence algorithms and corresponding types of membrane computing are carried out for band selection of WSVM ensemble model. As it is demonstrated in Section 3, MWOA is used here. To let the comparison impartially, all of algorithms are used to conduct coding for band combination of each sub-classifier, and the form is used with their standard mode.

4.1 Parameters setting for different algorithms

The optimization ability of WOA and other algorithms relies on some parameters setting to some extent. Table 1 lists the parameters setting of GSA [29], GWO [30], ALO [31] and the type of membrane computing in these comparative methods, as well as the parameters of WOA.

Among them, all of algorithms above are ended when the maximum number of object function evaluations reaches

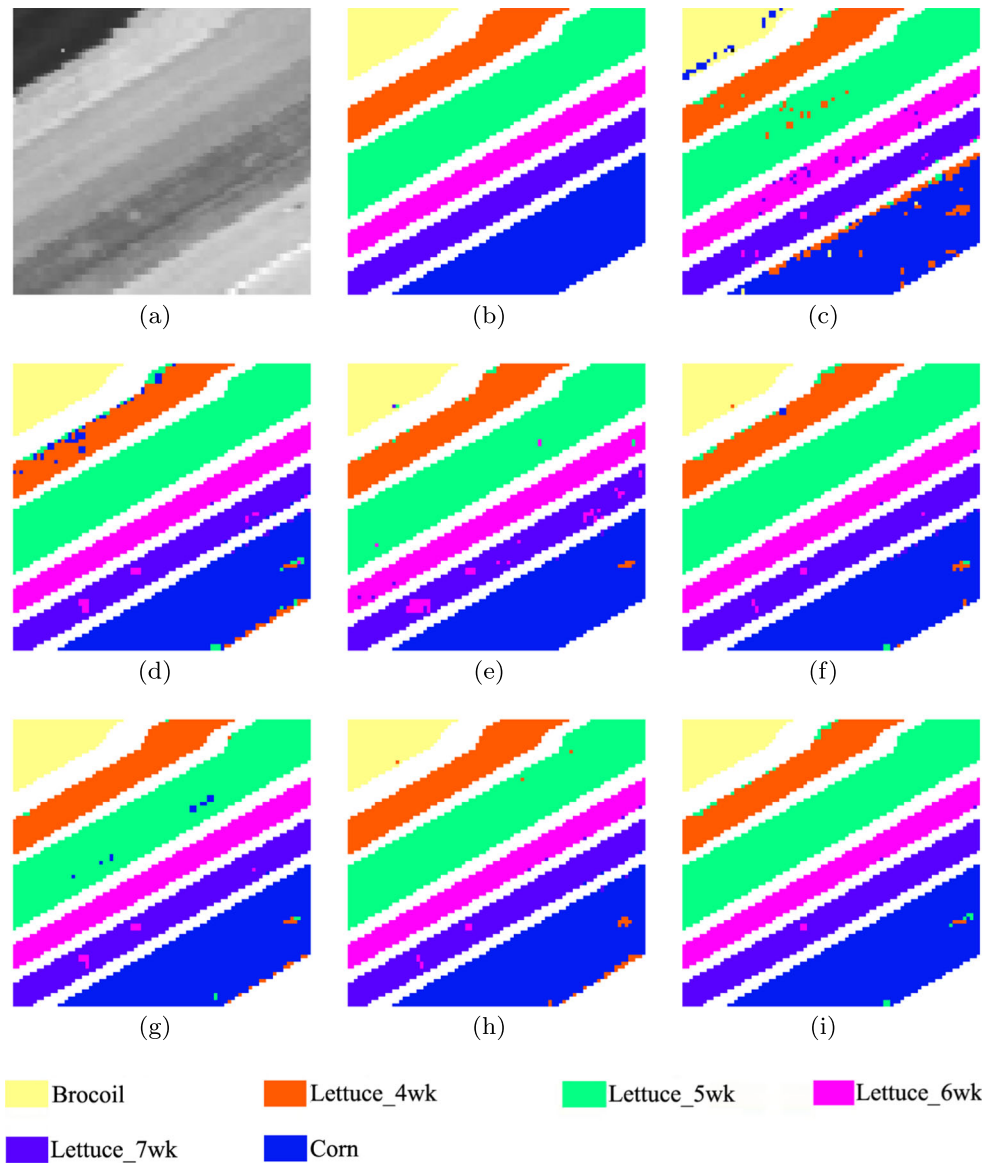

Fig. 4 Classification results for SalinasA: **a** original HSI **b** reference map **c** LS-SVM **d** WSVM **e** TCRC **f** RPN **g** DSVM **h** WOA **i** MWOA

2000, and 30 independent operations are executed for each algorithm. Meanwhile, some contrastive experimental results including illustrative examples and evaluating tables are listed in the section, which distinctly embody the advantages of the proposed WSVM ensemble model and MWOA. The primary task is to obtain the optimal band subset of HSI datasets, which is reflected by the value of (16), and the comprehensive performance of interpretation is reflected by overall classification accuracy (OA) with pixel-level.

4.2 Experiments for optimization ability

The above datasets are utilized in the subsection to prove the optimization ability of MWOA and the classification performance of WSVM ensemble model. Tables 2-3 show the property of classifier ensemble optimized by different algorithms and corresponding type of membrane computing, and Fiv, Acc, Ft and Time respectively denote the fitness value, classification accuracy, selected number of bands and CPU time in average after 30 independent operations.

As for the data in Tables 2-3, WOA has the stronger optimization ability compared with other algorithms, the fitness value is higher than 0.93 for SalinasA dataset, and only 3.9409s is cost to obtain a higher classification accuracy. As for measured datasets, the classification accuracy of WOA is higher than that of other two algorithms, but the selected number of bands is more than 19 that is still affiliated to scope of multi-spectral. Moreover, few of band information is selected as WOA combined with membrane computing, and CPU time is more than 60% decreased than before. More importantly, the classification accuracy reaches 92% for all datasets, which is higher than 98% for SalinasA dataset especially, and misclassification is obviously moderated by information integration of classifier ensemble. In brief, the optimization ability of WOA is the optimal, and the convergence speed is fast enough to

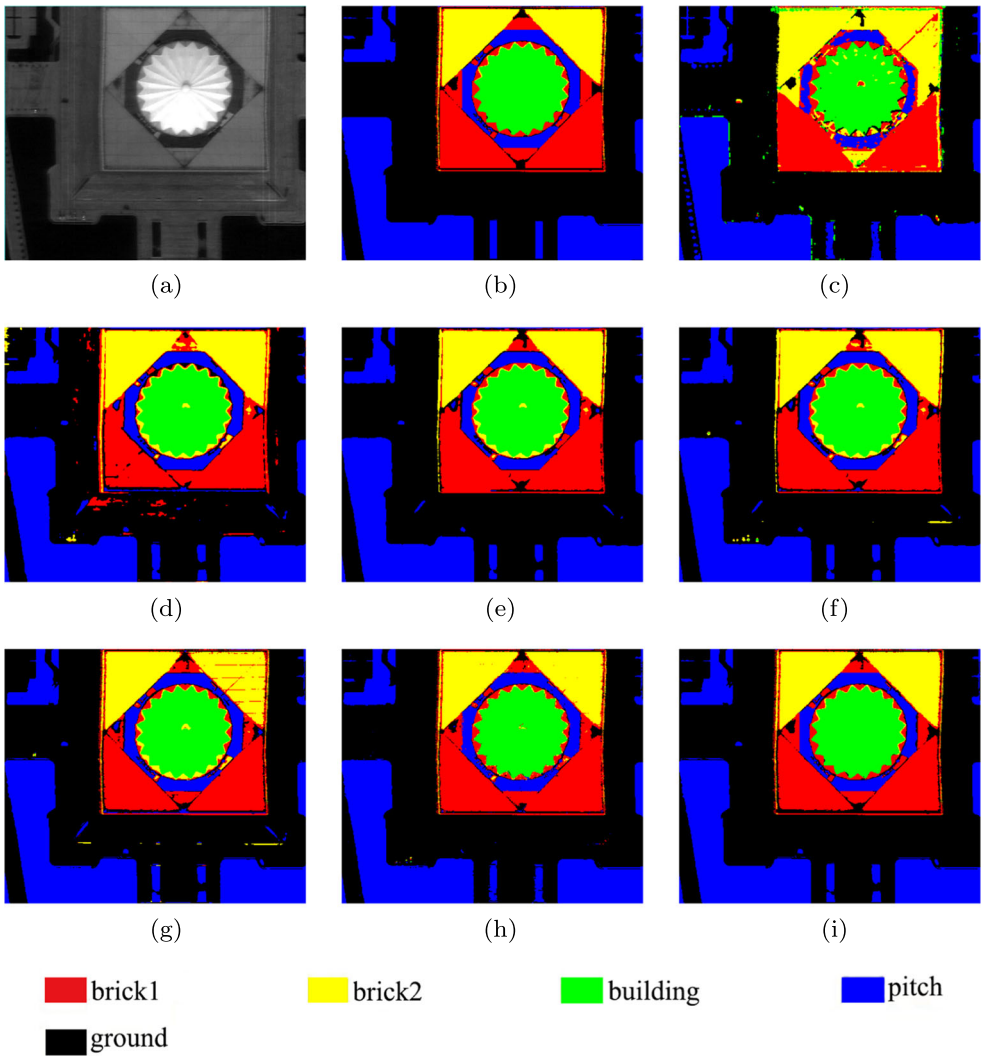

Fig. 5 Classification results for HSI1: **a** original HSI **b** reference map **c** LS-SVM **d** WSVM **e** TCRC **f** RPN **g** DSVM **h** WOA **i** MWOA

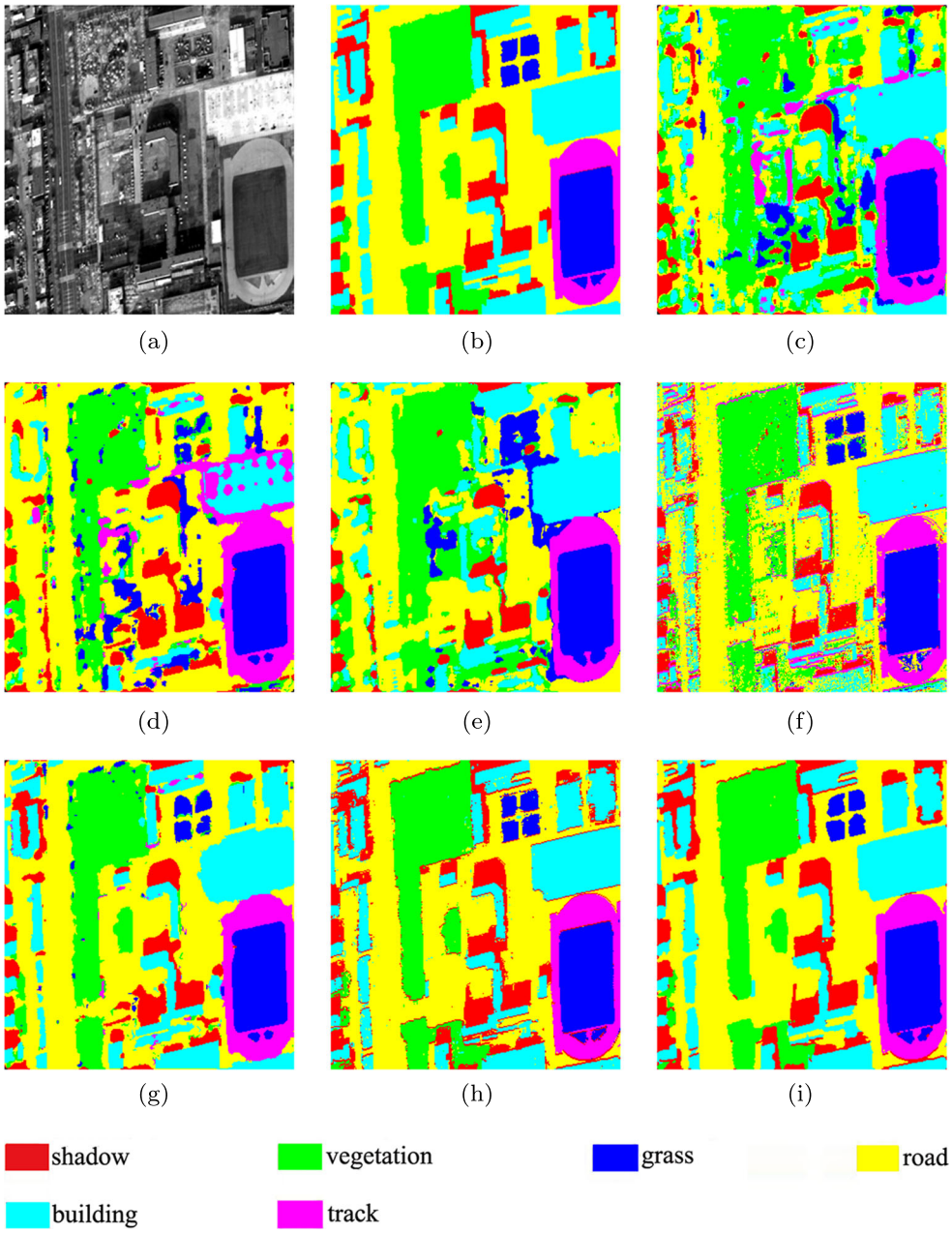

Fig. 6 Classification results for HSI2: **a** original HSI **b** reference map **c** LS-SVM **d** WSVM **e** TCRC **f** RPN **g** DSVM **h** WOA **i** MWOA

obtain higher accuracy with less band information by combining with membrane computing, WSVM ensemble model is suitable for HSI datasets to keep a good generalization ability, and the proposed technique is applicable for some practical work of band selection.

4.3 Experiments for pixel-level classification

In this subsection, four HSIs respectively named SalinasA and HSI1-HSI3 are used to conduct classification for each pixel of entire images. Moreover, some corresponding and newly proposed techniques such as, WSVM [23], local joint subspace (LS)-SVM [42], tangent collaborative representation classification (TCRC) (ensemble learning) [43], random patches network (RPN) (SVM for classification) [44] and deep SVM (DSVM) (SVM ensemble model via RBF kernel) [45] are also used to conduct a comparison here. The original and classified images have been listed on Figs. 4, 5, 6, and 7, and Table 4 outlines the OA of each HSI.

As for the classified images here, it is difficult to recognize different objects with similar feature values by using single classifier such as LS-SVM and WSVM, misclassification is obviously displayed in parts (c)-(d) of Figs. 4, 5, 6, and 7, some categories are leaked on the classified images, and the OA is lower than 72% for HSI2

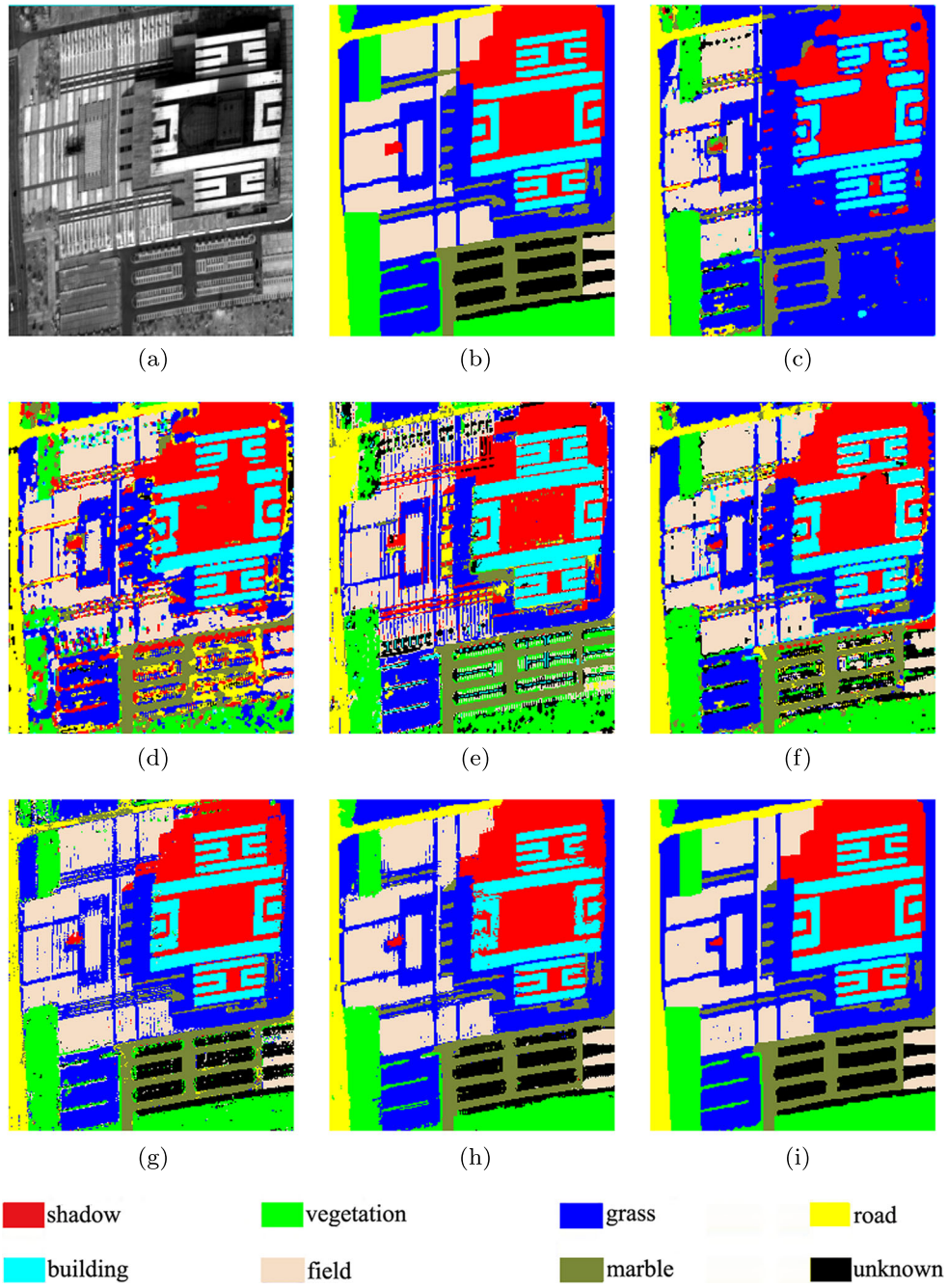

Fig. 7 Classification results for HSI3: **a** original HSI **b** reference map **c** LS-SVM **d** WSVM **e** TCRC **f** RPN **g** DSVM **h** WOA **i** MWOA

and HSI3 images. In part (e)-(f) of Figs. 4, 5, 6, and 7, the classification results are improved by the ensemble with a series of classifiers, but the discrimination ability of each sub-classifier is weak, and false category label is obtained for amount of samples by the voting strategy. Further, the OA is higher than 99% by using DSVM for SalinasA image, the classified image is coincided with the reference map to some extent, and it is superior to 83% for 3 measured images, but misclassification is still existed on the edge region because of spectra aliasing for the same type of SVM. For WSVM ensemble model, the homogeneity is improved by utilizing WSVM instead of traditional SVM, and the heterogeneity is simultaneously improved by different types of wavelet kernels compared with DSVM. The redundant band information is abandoned and remaining bands conduct more contribution for classification. More importantly, the OA is further improved as no duplicate band information inputs into sub-classifiers, and CPU time is reduced by using membrane computing with short coding length. The optimal band subset is obtained by the coding of MWOA as it is described in Section 3.2, and less than 10 bands are selected to compose the optimal band subset from

the HSIs with hundreds of bands, the optimal band subset with only 6 bands is obtained for HSI1 image, which is corresponding to the wavelengths of 1092.5nm, 1287.5nm, 1452.5nm, 1707.5nm, 1932.5nm and 2202.5nm. In sum, the proposed approach is specific and efficient that recognizes different objects from the entire HSIs with a reasonable computation efficiency, and it can be applied to conduct fast interpretation for HSIs.

5 Conclusion

In this paper, a band selection approach based on WSVM ensemble model and MWOA is proposed, and it is demonstrated that SVM with wavelet kernel is adapted to HSI datasets with high dimension and small samples, ensemble learning with different types of wavelet kernel is more appropriate to synthesize heterogeneity and homogeneity of each sub-classifier. In addition, WOA has excellent optimization ability, it is fast enough to obtain higher fitness value, and the coding length of each individual is decreased by membrane computing. Furthermore, experimental results are compared with some corresponding and newly proposed techniques for pixel-level classification, the OA has reached 93% for 4 HSIs by using the HSI classification technique in the paper, it is sufficiently discriminated for some objects with similar feature values and most of categories are recognized on the image. In general, WSVM is adapted to the feature values of band information collected from HSIs, misclassification is improved by ensemble learning and band selection, and the optimal band subset is obtained by MWOA. A good balance between computational efficiency and classification accuracy is maintained, which lets it more appropriate for a series of practical applications. In the future, it is prefer to fuse the spatial and spectral features combined with different types of sub-classifier and obtain the optimal feature subset on different scales.

Funding This work was funded by the National Natural Science Foundation of China under Grant No.41901296, 41925007, 41901285, the Key Laboratory for National Geographic Census and Monitoring, National Administration of Surveying, Mapping and Geoinformation under Grant No.2018NGCM06, and the Fundamental Research Funds for the Central Universities, China University of Geosciences (Wuhan) under Grant No.2642019046.

Ethics approval This article does not contain any studies with human participants or animals performed by any of the authors.

Declarations

Conflict of Interests Mingwei Wang and Ziqi Yan declare that they have no conflict of interest. Jianwei Luo declares that he has no conflict of interest. Zhiwei Ye declares that he has no conflict of interest. Peipei He declares that she has no conflict of interest.

References

1. Amigo JM, Babamoradi H, Elcoroaristizabal S (2015) Hyperspectral image analysis: A tutorial. Analytica Chimica Acta 896:34–51
2. Veraverbeke S, Dennison P, Gitas I et al (2018) Hyperspectral remote sensing of fire: State-of-the-art and future perspectives. Remote Sens Environ 216:105–121
3. Imani M, Ghassemian H (2020) An overview on spectral and spatial information fusion for hyperspectral image classification: Current trends and challenges. Inform Fusion 59:59–83
4. Tu B, Zhang X, Kang X et al (2019) Spatial density peak clustering for hyperspectral image classification with noisy labels. IEEE Trans Geosci Remote Sens 57:5085–5097
5. Hufnagl B, Lohninger H. (2020) A graph-based clustering method with special focus on hyperspectral imaging. Anal Chim Acta 1097:37–48
6. El Rahman S. A. (2015) Hyperspectral imaging classification using ISODATA algorithm: Big data challenge. International Conference on e-Learning. IEEE, 247–250
7. Sumarsono A, Du Q. (2016) Low-rank subspace representation for supervised and unsupervised classification of hyperspectral imagery. IEEE J Sel Top Appl Earth Obs Remote Sens 9:4188–4195
8. Liu C, Li J, He L. (2018) Superpixel-based semisupervised active learning for hyperspectral image classification. IEEE J Sel Top Appl Earth Obs Remote Sens 12:357–370
9. Haut JM, Paoletti ME, Plaza J et al (2018) Active learning with convolutional neural networks for hyperspectral image classification using a new Bayesian approach. IEEE Trans Geosci Remote Sens 56:6440–6461
10. Paoletti ME, Haut JM, Plaza J et al (2018) A new deep convolutional neural network for fast hyperspectral image classification. ISPRS J Photogramm Remote Sens 145:120–147
11. Sun G, Chen T, Su Y et al (2018) Internet traffic classification based on incremental support vector machines. Mobile Netw Appl 23:789–796
12. Liu R, Yang B, Zio E et al (2018) Artificial intelligence for fault diagnosis of rotating machinery: a review. Mech Syst Signal Process 108:33–47
13. Jain DK, Dubey SB, Choubey RK et al (2018) An approach for hyperspectral image classification by optimizing SVM using self organizing map. J Comput Sci 25:252–259
14. Liu L, Huang W, Liu B et al (2018) Semisupervised hyperspectral image classification via Laplacian least squares support vector machine in sum space and random sampling. IEEE J Sel Top Appl Earth Obs Remote Sens 11:4086–4100
15. Khodadadzadeh M, Ghamisi P, Contreras C et al (2018) Subspace multinomial logistic regression ensemble for classification of hyperspectral images. International Geoscience and Remote Sensing Symposium. IEEE, 5740–5743
16. Krawczyk B, Minku LL, Gama J et al (2017) Ensemble learning for data stream analysis: a survey. Inform Fusion 37:132–156
17. Sun J, Li H, Fujita H et al (2020) Class-imbalanced dynamic financial distress prediction based on adaboost-SVM ensemble combined with SMOTE and time weighting. Inform Fusion 54:128–144
18. Lin J, Chen H, Li S et al (2019) Accurate prediction of potential druggable proteins based on genetic algorithm and bagging-SVM ensemble classifier. Artif Intell Med 98:35–47
19. Feng W, Huang W, Bao W. (2019) Imbalanced hyperspectral image classification with an adaptive ensemble method based on SMOTE and rotation forest with differentiated sampling rates. IEEE Geosci Remote Sens Lett 16:1879–1883

20. Li Y, Xie T, Wang P et al (2018) Joint spectral-spatial hyperspectral image classification based on hierarchical subspace switch ensemble learning algorithm. Appl Intell 48:4128–4148
21. Falco N, Xia J, Kang X et al (2020) Supervised classification methods in hyperspectral imaging—recent advances. Data Handling in Science and Technology. Elsevier, 32, 247–279
22. Bilal M, Hanif M. S. (2019) High performance real-time pedestrian detection using light weight features and fast cascaded kernel SVM classification. J Signal Process Sys 91:117–129
23. Su H, Li X, Yang B et al (2018) Wavelet support vector machine-based prediction model of dam deformation. Mech Syst Signal Process 110:412–427
24. Sun W, Du Q. (2019) Hyperspectral band selection: A review. IEEE Geosci Remote Sens Mag 7:118–139
25. Wang Q, Zhang F, Li X. (2018) Optimal clustering framework for hyperspectral band selection. IEEE Trans Geosci Remote Sens 56:5910–5922
26. Nakamura RYM, Fonseca LMG, Dos Santos JA et al (2013) Nature-inspired framework for hyperspectral band selection. IEEE Trans Geosci Remote Sens 52:2126–2137
27. Su H, Du Q, Chen G et al (2014) Optimized hyperspectral band selection using particle swarm optimization. IEEE J Sel Top Appl Earth Obs Remote Sens 7:2659–2670
28. Ghosh A, Datta A, Ghosh S. (2013) Self-adaptive differential evolution for feature selection in hyperspectral image data. Appl Soft Comput 13:1969–1977
29. Tschannerl J, Ren J, Yuen P et al (2019) MIMR-DGSA: Unsupervised Hyperspectral band selection based on information theory and a modified discrete gravitational search algorithm. Inform Fusion 51:189–200
30. Medjahed SA, Saadi TA, Benyettou A et al (2016) Gray wolf optimizer for hyperspectral band selection. Appl Soft Comput 40:178–186
31. Wang M, Wu C, Wang L et al (2019) A feature selection approach for hyperspectral image based on modified ant lion optimizer. Knowl-Based Syst 168:39–48
32. Zhang X, Liu Z, Miao Q et al (2018) Bearing fault diagnosis using a whale optimization algorithm-optimized orthogonal matching pursuit with a combined time-frequency atom dictionary. Mech Syst Signal Process 107:29–42
33. Ala'M AZ, Faris H, Alqatawna J et al (2018) Evolving support vector machines using whale optimization algorithm for spam profiles detection on online social networks in different lingual contexts. Knowl-Based Syst 153:91–104
34. Chen H, Xu Y, Wang M et al (2019) A balanced whale optimization algorithm for constrained engineering design problems. Appl Math Model 71:45–59
35. Diaz-Pernil D, Gutierrez-Naranjo MA (2019) Membrane computing and image processing: A short survey. J Membr Comput 1:58–73
36. Dai H, Cao Z. (2017) A wavelet support vector machine-based neural network metamodel for structural reliability assessment. Computer-Aided Civ Inf Eng 32:344–357
37. Pathak RS, Singh A (2016) Mexican hat wavelet transform of distributions. Integral Transform Spec Funct 27:468–483
38. Li L, Liu P, Xing Y et al (2018) Time-frequency analysis of acoustic signals from a high-lift configuration with two wavelet functions. Appl Acoust 129:155–160
39. Tang B, Song T, Li F et al (2014) Fault diagnosis for a wind turbine transmission system based on manifold learning and Shannon wavelet support vector machine. Renew Energy 62:1–9
40. Cattani C. (2018) A review on Harmonic wavelets and their fractional extension. J Adv Eng Computat 2:224–238
41. Grana M, Veganzons M, Ayerdi B Resourse of SalinasA image. http://www.ehu.eus/ccwintco/index.php
42. Xiang P, Zhou H, Li H et al (2020) Hyperspectral anomaly detection by local joint subspace process and support vector machine. Int J Remote Sens 41:3798–3819
43. Su H, Yu Y, Du Q et al (2020) Ensemble learning for hyperspectral image classification using tangent collaborative representation. IEEE Trans Geosci Remote Sens 58:3778–3790
44. Xu Y, Du B, Zhang F et al (2018) Hyperspectral image classification via a random patches network. ISPRS J Photogramm Remote Sens 142:344–357
45. Okwuashi O, Ndehedehe CE (2020) Deep support vector machine for hyperspectral image classification. Pattern Recogn 103:107298

Mingwei Wang received his B.S. degree from Hubei Normal University, Huangshi, China, in 2011. He obtained the M.S. and Ph.D. degrees from Hubei University of Technology and Wuhan University, Wuhan, China, in 2015 and 2018, respectively. Since 2018, he is a professional researcher in the Institute of Geological Survey, China University of Geosciences, Wuhan, China. His major research interests include hyperspectral image processing, swarm intelligence algorithm and deep learning model.

Ziqi Yan received her B.S. degree from China University of Geosciences, Wuhan, in 2017. Since 2017, she is working toward the M.S. degree at China University of Geosciences, Wuhan, China. Her major research interests include hydrologic model research based on GIS, remote sensing image processing and swarm intelligence algorithm.

Jianwei Luo received his B.S. degree from Wuhan University of Technology, Wuhan, China, in 2005. Since 2012, he is a senior engineer in the Hubei Cancer Hospital, Huazhong University of Science and Technology, Wuhan, China. His major research interests include hyperspectral imaging theory, machine learning model and computer application technology.

Zhiwei Ye received his Ph.D. degree in Photogrammetry from Wuhan University, Wuhan, China in 2006. He is a professor in school of computer science, Hubei University of technology, Wuhan, China. He has published more than thirty papers in the area of digital image processing and swarm intelligence algorithm. His research interests include image analysis, pattern recognition and data mining.

Peipei He received her B.S. degree from Information Engineering University, Zhengzhou, China, in 2009; and received the M.S. and Ph.D degrees from Wuhan University, Wuhan China, in 2012 and 2015, respectively. She is currently a lecturer at the College of Surveying and Geo-Informatics, North China University of Water Resources and Electric Power. Her research interests include multi-source data registration, precise classification of laser scanning data.

Affiliations

Mingwei Wang[1,2] · Ziqi Yan[1] · Jianwei Luo[3] · Zhiwei Ye[4] · Peipei He[5]

Mingwei Wang
wangmingwei@cug.edu.cn

Ziqi Yan
Yanziqi@cug.edu.cn

Zhiwei Ye
weizhiye121@163.com

Peipei He
hepei@ncwu.edu.cn

1 Institute of Geological Survey, China University of Geosciences, Wuhan, 430074, People's Republic of China

2 Key Laboratory for National Geographic Census and Monitoring, National Administration of Surveying, Mapping and Geoinformation, Wuhan, 430079, People's Republic of China

3 Hubei Cancer Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, 430079, People's Republic of China

4 School of Computer Science, Hubei University of Technology, Wuhan, 430068, People's Republic of China

5 College of Surveying and Geo-Informatics, North China University of Water Resources and Electric Power, Zhengzhou, 450045, People's Republic of China