A Cooperative Framework Based on Active and Semi-supervised Learning for Sea Ice Classification using EO-1 Hyperion Data*

Yanling HAN,†1 Yao ZHAO,†1 Yun ZHANG,†1 Jing WANG,1 Shuhu YANG,†1 Zhonghua HONG,†1 and Shouqi CAO(2)

1College of Information Technology, Shanghai Ocean University, Shanghai 201306, China
2College of Engineering Science and Technology, Shanghai Ocean University, Shanghai 201306, China

In the classification of remote-sensing sea ice images, labeled samples are difficult to acquire. To adequately utilize the massive number of unlabeled samples, which contain abundant information, we propose a cooperative framework based on active learning (AL) and semi-supervised learning (SSL) for sea ice image classification. We acquire the most valuable samples using AL and make full use of the abundant information contained in the unlabeled samples using SSL, and then conduct a label consistency verification procedure to further ensure the quality of the pseudo-labeled samples obtained through cooperation between AL and SSL. In the AL part, we adopt a sampling strategy that integrates uncertainty and diversity criteria to acquire the most valuable samples to label. In the SSL part, we utilize the SSL sampling strategy to choose the unlabeled samples with the most information and little redundancy, and use the transductive support vector machine (TSVM) as the classification model. The cooperation between AL and SSL ensures the accuracy of the pseudo-labeled samples through a consistency verification procedure. We conduct comparative experiments using the method proposed and other methods on two hyperspectral images obtained from the Earth Observation Satellite 1 (EO-1). The proposed method achieves the highest classification accuracy for both datasets and can be effectively applied to sea ice classification.

Key Words: Active Learning, Cooperative Training, Remote-sensing Image, Sea Ice Classification, Semi-supervised Learning

Nomenclature

AL: active learning
SSL: semi-supervised learning
SVM: support vector machine
TSVM: transductive support vector machine
EO-1: Earth Observation Satellite 1

1. Introduction

Sea ice disasters produce huge losses in maritime transportation, shipping, marine fisheries, and marine oil and gas infrastructure in polar and high-latitude regions. Therefore, research on disaster prevention and the classification of sea ice has received increasing attention.1) Remote-sensing technology is an efficient method for the detection of sea ice. Compared with the traditional methods of detecting sea ice, remote-sensing technology can rapidly acquire information from a large area and has been widely used in sea ice detection.2) During recent years, the data sources commonly used include synthetic aperture radar,3–5) multispectral satellite images with medium or high spatial resolution (e.g., images obtained when using a imaging spectroradiometer),6–8) Landsat,9) and hyperspectral images.10) Of these, the hyperspectral remote sensing-images contain more spectral information, and an abundant amount of spatial information and have a higher capability for target recognition, which can greatly improve the accuracy of target detection.

In hyperspectral remote-sensing sea ice image classification, due to the environmental limitations and conditions, the annotation of remote-sensing images is more time consuming and laborious, which results in a small number of available labeled sea ice samples. However, a large number of unlabeled samples have not been fully utilized. To solve the problem of poor classification accuracy due to the small number of labeled samples, active learning11) and semi-supervised learning12) are being applied to remote-sensing image classification.

Active learning (AL) is a process similar to that of human-computer interaction. Using certain query criteria, the AL model can repeatedly search for unlabeled samples containing a large amount of information. It then adds them to the training sample set via expert manual annotation using a supervised classification approach. This makes it possible to avoid adding samples with little information content. In doing so, an AL model can greatly reduce the cost of labeling and improve the quality of the training samples. At present, investigators have proposed many relevant AL methods for remote sensing image classification, such as margin sampling (MS),13) the best versus second-best (BvSB) strategy,14) the multiclass-level uncertainty (MCLU) method,15) entropy-based uncertainty sampling,16) and so on. MS is an uncertainty sampling method that takes advantage of the dis-
tance from a sample to a hyperplane to select unlabeled samples. The unlabeled samples in these regions are more likely to be support vectors, so they are uncertain and have more information than other unlabeled samples. However, this method only works for binary classification problems. The MCLU method is an uncertainty sampling method that is proposed for dealing with multiclass classification problems. In this method, the difference between the first and second largest distance to a hyperplane is calculated for each unlabeled sample. Then, a certain number of unlabeled samples with the smallest differences are selected to be manually labeled. Since the MCLU method merely selects the samples with the largest uncertainty, there are similar samples with redundant information among them. In recent years, investigators have developed a more effective AL sampling strategy that involves selecting the most informative samples according to the clustering hypothesis. Therefore, a sampling process that is based on the diversity of the sampling is introduced to reduce the redundancy of the manually labeled samples. This AL method with a clustering conjecture is divided into two parts: an uncertainty sampling section and a differential selection section. In the uncertainty sampling section, \( m (m > 2) \) samples are selected using the uncertainty criterion, and then \( h (1 < h < m) \) samples, which are as diverse as possible, are selected from the \( m \) samples using the diversity criterion. Demir et al. studied several AL methods combining differentiated criteria and uncertainty criteria to classify remote-sensing images. Clustering-based diversity (CBD) is a sampling method for measuring the diversity among unlabeled samples. It uses a clustering method based on standard \( k \)-means clustering, which clusters similar, but unlabeled, samples into the same cluster in the feature space. Unlike the CBD method, the ECBD method divides the samples into different clusters based on the \( k \)-means clustering method in the kernel feature space. In the entropy query by bagging (EQB) method, the original training set is divided into \( k \) training sets. Then, each training set is used for training the model, and the unlabeled samples are predicted. Each sample is predicted to have \( k \) labels. In this method, the entropy value is used to measure the amount of information in the predicted tags, and the samples with the maximum entropy values are selected. Patra and Bruzzone proposed a new AL algorithm using a self-organizing map (SOM) to query differential samples and incorporated the uncertainty criteria. Shi et al. integrated a relearning model into the AL framework, aiming to accelerate the convergence of the AL algorithm and further reduce the labeling cost: two spatial–spectral uncertainty criteria to optimize the procedure for selecting new training samples are also proposed.

Compared to AL, semi-supervised learning (SSL) saves time and cost by using a small number of labeled samples to directly update the classifier; that is, a large number of unlabeled samples without any manual labeling is used. The SSL approach mines a large amount of available information from the unlabeled samples. The hypothesis or internal relationships in the distribution of the samples is then used to expand the samples after marking them. This improves the performance of the classifier. Dópido et al. proposed a self-training technique for the classification of hyperspectral images. The method first selects confident candidate unlabeled samples based on the spatial and spectral information. Then, the most informative samples among the candidate samples are selected to be included in the training set by adopting an AL method. Camps-Valls et al. proposed the graph-based semi-supervised method based on local and global consistency. In this method, every point iteratively spreads its label information to its neighbors until a global stable state is achieved. These methods combine spatial, spectral, and spatial–spectral information to improve the classification robustness and accuracy. Vapnik and Stin proposed using the transductive support vector machine (TSVM) method, which is used to estimate the linear prediction function of the class labels. It is an extension of the standard support vector machine with unlabeled samples and is based on an iterative algorithm. During the initial iterations, the standard SVM is exploited to obtain the initial hyperplane. The TSVM method utilizes both the labeled and unlabeled samples in the training process. Zhou and Li proposed an SSL method of learning by disagreement, which uses the differences among multiple learning machines to improve SSL performance. This method effectively reduces the time cost and improves the generalization ability of the learning machine. Bruzzone et al. proposed a progressive semi-supervised support vector machine (PS\( ^* \)VM) to select samples between, and close to, the classification surfaces to improve the performance of the semi-supervised classifier. Aydav and Minz proposed a modified self-training approach that selects informative and confident samples using a pseudo-validation set to improve the scope of self-training. The granulation-based self-training method proposed selects correct, diverse and informative samples that effectively improve classifier performance.

Both the AL and SSL approaches start from a small number of labeled samples and use the unlabeled samples to iteratively improve the performance of the classification model. Recently, the AL and SSL approaches have been combined into a collaborative method. The combination of AL and SSL can effectively reduce the human labeling efforts while obtaining a reliable classification performance with unlabeled samples. Some combination methods have good performance. Zhang and Crawford proposed a semi-supervised multiple metric learning method for feature extraction and combined it with AL into a unique framework. The metric learning approach proposed learns distinct projection matrices jointly, and each metric is assigned to one type of feature. Moreover, the regularization proposed helps avoid overfitting by taking advantage of the unlabeled data information. Finally, different types of features are projected into a common feature space where AL is applied in conjunction with \( k \)-nearest neighbor classification. Wan et al. proposed a collaborative active and semi-supervised learning (CASSL) approach. The method combines AL and SSL to invoke a collaborative labeling process using both human experts and classifiers. Specifically, the AL-based pseudo-label ver-
The classification procedure is used to improve the pseudo-labeling accuracy to facilitate SSL. Leng et al.\textsuperscript{32}) proposed an active semi-supervised SVM algorithm. This method first performs $M$ iterations of AL and records the predicted label for the unlabeled sample in each iteration. It then calculates the $M$-th distance between the samples and the classification hyperplane. Next, the samples with a label changing rate (LCR) equal to zero are selected as the reliable samples. The class central sample is calculated based on the median distance (MD) from the sample to the classification hyperplane among the selected samples. Finally, the class central sample and its corresponding prediction labels are added to the pseudo-labeled sample set to train the semi-supervised classifier. The above steps are iteratively repeated to update the classifier to predict the test sample and obtain the final classification result. In Zhou and Prasad,\textsuperscript{33}) AL is combined with SSL to classify hyperspectral images using the morphological component analysis (MCA) method. First, the original hyperspectral data are decomposed into morphological components using MCA. In each feature domain, AL is combined with SSL to amplify the training datasets based on superpixels; and finally, decision fusion is used to integrate the predictions from the two components. Liu et al.\textsuperscript{34)} presented a new superpixel-based semi-supervised AL method that introduces the concept of spatial adaptivity into pseudo-labeling to improve machine generalization.

To further improve the performance of the classifier and reduce the manual annotation workload, this paper proposes a new cooperative classification framework. This is done by combining an AL sampling strategy and a TSVM SSL technology for sea ice image classification called “cooperative framework based on active learning and a transductive support vector machine” (CFATSVM). It takes full advantage of the capabilities of AL and SSL, and combines the two methods organically. In the AL part, the uncertainty criteria based on the BvSB method and diversity criteria based on the SOM and ECBD methods are combined to query the samples with abundant information from the low-density regions of the samples. The significant diversity in the kernel space, where a support vector machine (SVM) separates the hyperplane, contributes to improving the performance of the SVM classifier. In the SSL part, a large number of unlabeled samples are queried from samples inside the margin of the current classifier, which contains an abundance of information and diversity. The TSVM method is used to train the SSL classifier to enable the semi-supervised classification process. We simultaneously conduct label consistency verification by comparing the results obtained from the SVM and TSVM. This is done to ensure pseudo-label reliability (i.e., labels are assigned to unlabeled samples by a classifier) for the unlabeled samples selected. The experimental results obtained from two hyperspectral remote-sensing datasets show that the method proposed has better classification performance when compared to the performance of AL and SSL using some of the classic cooperative methods.

The remainder of this paper is arranged as follows. Section 2 introduces the AL and SSL sampling strategies used in the CFATSVM framework, and details the framework and implementation of the CFATSVM algorithm. The experimental results and analysis are presented in Section 3. Section 4 concludes this paper.

2. CFATSVM Framework Proposed

By combining the AL and SSL models, the algorithm can label the samples using human experts as the traditional AL algorithm does. Then, using a classifier in the SSL process, a promising solution for remote-sensing image classification of sea ice can be obtained using a small number of samples. In this paper, we propose the CFATSVM classification framework. This framework can achieve better classification performance by using cooperative labeling based on the AL and SSL algorithms and a few labeled samples. Before introducing the CFATSVM framework, we first introduce the AL sampling strategy and SSL technology used in this framework.

2.1. AL strategy in the CFATSVM algorithm

The sampling strategy is a key part of AL. AL is used to select the most informative samples according to the sampling strategy. It then adds them into the labeled sample set through manual expert labeling, and trains the new classification model with the updated labeled sample set. The AL strategy in the CFATSVM algorithm integrates the uncertainty criterion based on the BvSB algorithm and the diversity criterion based on the SOM and ECBD algorithms to form a new query strategy, which is called Q1.

In the AL of the CFATSVM method, the BvSB criterion is used to consider the difference between the probability values of the two classes having the highest estimated probability value as a measure of uncertainty. Using such a measure is a more direct way of estimating the confusion about class membership from a classification standpoint.

Then, the SOM neural network can select the informative samples from the low-density regions of the feature space by exploiting the topological properties. When the training phase converges, we can calculate the average neighbor distance of each neuron in the output layer to its neighboring neurons using the corresponding weight vectors.\textsuperscript{19)} The average neighbor distance of neuron $k$, denoted as $\overline{w}_k$, is computed as follows:

$$\overline{w}_k = \frac{1}{|N_f|} \sum_{i \in N_f} \| w_k - w_i \|^2$$

where, $N_f^r$ represents the set of neurons in the $r$th-order neighbor system of neuron $k$ in the map.

Next, the ECBD clustering algorithm works by applying kernel $k$-means clustering in the kernel space where the SVM separating hyperplane operates.\textsuperscript{10,15)} Because of the kernel mapping, the set of the most diverse samples in the original space may not be the most diverse in the kernel space. The samples are mapped to a high-dimensional feature space, as it is possible to eliminate the linear inseparability in the original space by solving linear separable problems there. The samples are then divided into $h$ clusters. The ECBD al-
algorithm selects the samples in each cluster center as the representative cluster samples.

The AL strategy can query the samples with abundant information in the low-density area of samples through the BvSB and SOM techniques. Then, the ECBD algorithm is used to select the representative cluster samples with diversity in the kernel space where the SVM separating hyperplane operates. The AL algorithm, Q1, is summarized as follows.

Algorithm 1: AL algorithm, Q1
Inputs:
Classifier: C
Unlabeled sample pool: U
Query size: h2, h1, q1
Output:
Unlabeled sample sets for human expert labeling: Uq1
Start:
1: Train the SOM neural network using all of the available labeled samples.
2: Compute the average neighbor distance of each neuron using Eq. (1).
3: Select h1 samples using classifier C and the BvSB criterion from U, which have the lowest certainty values and are mapped onto distinct neurons in the SOM.
4: Select the h1 (h1 < h2) samples from the h2 samples that correspond to the SOM mapping neurons having the highest average neighbor distances.
5: Select q1 (1 < q1 < h1) samples with diversity from the h1 samples for manual labeling according to the ECBD criterion.
End.

2.2 SSL strategy in the CFATSVM algorithm
SSL is a method to improve the performance of classifiers using a large number of unlabeled samples. TSVM is an effective SSL technology. The concept of a TSVM was proposed by Vapnik and Sterin.\(^{24}\) The main idea is to build a classifier from a large number of unlabeled samples. TSVM is an effective SSL technology. The concept of a TSVM was proposed by Vapnik and Sterin.\(^{24}\) The main idea is to build a classifier from a large number of unlabeled samples, and also adds the implicit distribution information of the unlabeled samples to the standard training process of the SVM algorithm. Algorithm performance is greatly improved, and it also reduces reliance of the standard SVM on a large number of labeled samples. The following is a brief introduction to the principles of the TSVM.

Set \(l\) consists of independent and identically distributed labeled samples

\[
\{x_i, y_i\}_{i=1}^l, \quad y_i \in \{-1, +1\}
\]  

where, \(x_i\) represents the labeled samples, \(y_i\) represents the labels of \(x_i\), and \(k\) represents the unlabeled samples

\[
\{x_i^+\}_{i=1}^k
\]  

In general, the learning process of a TSVM can be formulated as the following optimization problem:

\[
\begin{align*}
\min \left( y_1^+, \ldots, y_l^+, w, b, \xi_1, \ldots, \xi_l, \xi_1^+, \ldots, \xi_l^+ \right) \\
\frac{1}{2} \|w\|^2 + C \sum_{i=1}^l \xi_i + C^* \sum_{j=1}^k \xi_j^+ \\
\text{subject to:} \quad \langle w, x_i \rangle + b \geq 1 - \xi_i, \quad \xi_i \geq 0 \\
\langle w, x_j^+ \rangle + b \geq 1 - \xi_j^+, \quad \xi_j^+ \geq 0
\end{align*}
\]  

where, \(w\) is the normal vector to the hyperplane, \(b\) is the offset to the hyperplane, \(l\) is the number of labeled samples, \(x_i\), and \(k\) is the number of unlabeled samples, \(x_j^+\), for transductive learning. \(C\) and \(C^*\) are user-specified parameters: \(C\) is called the ‘effect factor of the labeled samples,’ and \(C^*\) is called the ‘effect factor of the unlabeled samples.’ \(\xi_i\) and \(\xi_j^+\) are the slack variables for both training and unlabeled samples. To facilitate solving the quadratic programming problem of Eq. (4), similar to the standard SVM, a kernel function and Lagrange multiplier are also introduced. Then, the quadratic optimization problem of Eq. (4) is transformed into:

\[
\begin{align*}
\min \frac{1}{2} \left( \sum_{i=1}^l \sum_{j=1}^l K(x_i, x_j) y_i y_j \alpha_i \alpha_j \right) \\
+ 2 \sum_{i=1}^l \sum_{j=1}^k K(x_i, x_j^+) y_i y_j \alpha_i \alpha_j^+ \\
+ \sum_{i=1}^k \sum_{j=1}^l K(x_i^+, x_j) y_i y_j^+ \alpha_i^+ \alpha_j^+ - \left( \sum_{i=1}^l \alpha_i + \sum_{j=1}^k \alpha_j^+ \right)
\end{align*}
\]  

subject to:

\[
0 \leq \alpha_i \leq C, \quad 1 \leq i \leq l \\
0 \leq \alpha_j^+ \leq C, \quad 1 \leq j \leq k \\
\sum_{i=1}^l y_i \alpha_i + \sum_{j=1}^k y_j^+ \alpha_j^+ = 0
\]  

where, \(\alpha^+\) is the Lagrange multiplier of the labeled samples and \(\alpha^\ast\) is the Lagrange multiplier of the unlabeled samples. \(K(x_i, x_j)\) is the radial basis function (RBF) kernel:

\[
K(x_i, x_j) = \exp \left( -\gamma \| x_i - x_j \|^2 \right)
\]  

where, \(\gamma\) is a Gaussian kernel parameter that can be optimized by grid optimization in the SVM. The final decision function is as follows:

\[
f(x) = \text{sgn} \left[ \sum_{i=1}^l \alpha_i y_i K(x, x_i) + \sum_{j=1}^k \alpha_j^+ y_j^+ K(x, x_j^+) + b \right]
\]  

Due to the large number of unlabeled samples, we utilize the idea of AL to select the samples with a large amount of information and improve the process of selecting the unlabeled samples of the TSVM. The MCLU method selects the most informative samples according to a multiclass confidence value, which is defined on the basis of the discrimination function of the binary SVMs.\(^{14}\) The mathematical model can be expressed as:

\[
C(x) = f_1(x) - f_2(x)
\]  

where, \(f_1(x)\) and \(f_2(x)\) are the first and second highest output scores of the binary SVMs, respectively. Similar to the binary
case, querying the sample that minimizes \( C(x) \) results in the selection of the sample that is closest to the boundary between the two most likely classes. The unlabeled samples are selected considering the multiclass confidence measure defined in Eq. (9) combined with a diversity criterion. In this way, both labeled and unlabeled samples are considered. Let us define the set of samples, \( H \), as follows:

\[
H = \{ x \mid x \in U, 0 < C(x) < 2 \} \tag{9}
\]

\( H \) contains all of the samples inside the margin. We define the set of candidate unlabeled samples, \( J \), in \( H \) with \( C(x) \) closer to 2. To further reduce the number of unlabeled samples, we select \( t \) \((t < J)\) samples among the candidate samples, \( J \). In the first step, we initialize the set of selected samples, \( U_{q1} \), with sample \( x_{up} = \text{argmax}_{x \in J} C(x) \). Then, we incrementally include in \( U_{q1} \) the samples in \( J \) that minimize the similarity with the closest sample already included in the set. To avoid an exhaustive search, we incorporate a diversity measure that considers the angles between the hyperplanes induced. Then, we select \( q2 \) diverse samples among the \( t \) candidates. The similarity between two samples is computed considering kernel cosine-angular similarity,

\[
K^a(x, x_i) = \frac{K(x, x_i)}{\sqrt{K(x, x)K(x, x_i)}} \tag{10}
\]

where, \( K^a(x, x_i) \) is also an RBF kernel, \( x \) and \( x_i \) are the samples selected from \( H \), and \( x \) is a temporary cluster center sample, \( x_i \) is the \( i \)-th sample. The specific process of the sampling strategy, \( Q2 \) for the TSVM unlabeled training samples is shown in Algorithm 2.

**Algorithm 2:** The SSL sampling strategy, \( Q2 \)

**Inputs:**
- Classifier: \( F \)
- Unlabeled set: \( U \)
- Query size: \( t, q2 \)

**Output:** TSVM unlabeled training samples, \( U_{q2} \)

**Start:**
1: According to classifier \( F \), Eq. (8) and Eq. (9), select \( t \) samples with a higher \( C(x) \) value inside the margin from \( U \) to form the set \( J \).
2: According to Eq. (8), select the sample with the maximum value of \( C(x) \) from \( J \) as \( x_{up} \).
3: Take \( x_{up} \) as the initial sample in the TSVM unlabeled training sample set, \( U_{q1} \). \( (U_{q1} = \{x_{up}\}) \).
4: **Repeat.**
5: Select a sample with the minimum similarity to the samples in \( U_{q1} \),

\[
x_{up} = \text{argmin}_{x \in J \setminus U_{q1}} \{\max_{x_i \in U_{q1}} K^a(x, x_i)\}.
\]

6: Add the updated \( x_{up} \) to \( U_{q2} \).
7: **Until** \( |U_{q2}| = q2 \).

**End.**

2.3. Detailed steps of the CFATSVM algorithm

The framework and pseudo-code of the CFATSVM algorithm proposed are illustrated in Fig. 1 and Algorithm 3, given an initial labeled sample set, \( L \), unlabeled sample set, \( U \), empty pseudo-labeled sample set, \( P \), and empty unlabeled training sample set, \( S \).

At the beginning, the TSVM classifier is trained jointly using the labeled sample set, \( L \), pseudo-labeled dataset, \( P \), and TSVM unlabeled training sample set, \( S \). Since \( S \) and \( P \) are empty, this step is actually a standard SVM training process. According to the query strategy, \( Q2 \), described in Algorithm 2, \( q2 \) unlabeled samples, \( U_{q2} \), having a certain credibility and diversity are selected using the classifier obtained and added into the unlabeled training sample set, \( S \). The updated TSVM classifier is trained using sets \( L \) and \( P \) and the updated sample set, \( S \). When the TSVM algorithm uses an unlabeled sample set, \( S \), for training, it needs to set the value of \( N^+ \), which is the number of positive samples in set \( S \). We use the current classifier to predict labels for samples in \( S \) and then count the number of positive labels to set the value of \( N^+ \).

Then, the samples in \( S \) are added to the candidates for the pseudo-labeled sample set, \( CP \), and \( S \) and \( U \) are updated. The TSVM classifier is used to predict the samples in \( CP \) to obtain the classification results, \( label_T \).

Then, \( q1 \) unlabeled samples, \( U_{q1} \), are selected according to the TSVM classifier and the query strategy, \( Q1 \), described in Algorithm 1, and are labeled by human experts. These \( q1 \) newly labeled samples are removed from \( U \) and added into \( L \). The SVM classifier is trained using the updated \( L \) dataset, which is then applied to classify the samples in \( CP \) to obtain the classification results, \( label_S \).

Next, the pseudo-label verification procedure is carried out based on the classification results, \( label_T \) and \( label_S \). The samples with the same classification results comparing \( label_T \) with \( label_S \) are added to the pseudo-labeled dataset, \( P \), with the corresponding pseudo-label. Dataset \( P \) is used to train a new TSVM classifier. The pseudo-label verification procedure is used to improve the quality of the pseudo-labeled samples for training.

Finally, \( q2 \) samples are selected from \( U \) and added into \( S \) using the TSVM classifier and query strategy, \( Q2 \), and are used to prepare for training the TSVM classifier in the next iteration. The \( q2 \) samples selected are removed from \( U \) and the algorithm will terminates once \( U \) is empty.

**Algorithm 3:** The CFATSVM algorithm

**Inputs:**
- Initially labeled sample set: \( L \)
- Unlabeled sample pool: \( U \)
- Pseudo-labeled dataset: \( P = \emptyset \)
- TSVM unlabeled training sample set: \( S = \emptyset \)
- Candidates for the pseudo-labeled sample set: \( CP = \emptyset \)

**Start:**
1: **Repeat.**
2: Update \( LP = L \cup P \), then train the TSVM classifier using \( LP \) and \( S \).
3: If \( S \) is empty, execute Step 8; otherwise, execute Step 4.
4: Update \( CP = CP \cup S \), then classify \( CP \) using the TSVM classifier and get the classification result labels, \( label_S \).
5: Select \( q1 \) samples, \( U_{q1} \), from \( U \) based on the TSVM
3. Experimental Results and Discussion

To show the effectiveness of the CFATSVM algorithm proposed, the experiments were performed using two hyperspectral sea ice images obtained from EO-1. The sea ice data are introduced in Section 3.1. Section 3.2 describes the experimental setup for the CFATSVM approach and other methods for comparison. Then, the detailed classification results obtained for the two sea ice images are presented: the Baffin Bay dataset results in Section 3.3 and the Bohai Bay dataset results in Section 3.4. Moreover, we perform a parameter sensitivity analysis for the CFATSVM method in Section 3.5.

3.1. Description of the datasets

The experiment used two hyperspectral sea ice remote-sensing images acquired by the Hyperion sensor mounted on EO-1. The scenes selected were of a partial area of Baffin Bay in Greenland and Bohai Bay in China. The hyperspectral images have a total of 242 bands and a spatial resolution of 30 m. The number of bands is initially reduced to 176 by removing the bands with low signal-to-noise ratio and water absorption.

Due to the unique environment of sea ice, it is difficult to obtain the ground truth. When no real surface information is available for comparison, a remote-sensing image with a higher spatial resolution from the same time and location can be taken as the reference image, and the classification results are used as the real distribution situation of the ground truth.

Baffin Bay dataset: The hyperspectral image in this dataset was acquired on April 12, 2014. The image size is 2395 × 1769 pixels. According to the spectral reflectance, the data are divided into three categories: white ice, gray ice, and sea water because the different types of sea ice can be discriminated by their spectral reflectance. The available labeled samples (1678 samples) were collected by Landsat-8 image interpretation as shown in Fig. 2. The samples are selected randomly as the initial training sample set, \( L \), and the rest is taken as the unlabeled sample set, \( U \). Finally, Landsat-8 data with a spatial resolution of 15 m are used for label verification at the same time using the same scene. The sea ice type and number of related samples are shown in Table 1.
Table 1. Number of samples for each class in the initial training set ($L$) and in the unlabeled pool ($U$) for the Baffin Bay dataset.

| Class          | $L$ | $U$ | Total |
|----------------|-----|-----|-------|
| White ice      | 3   | 540 | 543   |
| Gray ice       | 3   | 536 | 539   |
| Sea water      | 3   | 593 | 596   |
| Total          | 9   | 1669| 1678  |

Bohai Bay dataset: The hyperspectral image in this dataset was acquired on January 23, 2008. The image size is $823 \times 446$ pixels. According to the spectral reflectance, the data are divided into three categories: white ice, white-gray ice, and gray ice. The number of available labeled samples is 1800. From the total sample set, nine are selected randomly as the initial training sample set, $L$. Finally, data from Landsat-7 having a spatial resolution of 15 m is used for label verification at the same time using the same scene. Because the scan line corrector (SLC) in the enhanced thematic mapper plus (ETM+) instrument of Landsat-7 failed on May 31, 2003, approximately 22% of the data in the Landsat-7 scene were missing. Therefore, some strips in the Landsat-7 image, part of the original data, are missing. To effectively evaluate the data detection results, we applied mask processing to the hyperspectral data to make it consistent with the Landsat-7 data, as shown in Fig. 3. The sea ice type and the number of related samples are shown in Table 2.

| Class          | $L$ | $U$ | Total |
|----------------|-----|-----|-------|
| White ice      | 3   | 597 | 600   |
| White-gray ice | 3   | 597 | 600   |
| Sea ice        | 3   | 597 | 600   |
| Total          | 9   | 1791| 1800  |

The number of positive class samples, $N^+$, required in the TSVM training process is set according to the prediction result of the classifier using the unlabeled training sample set, $S$.

In the CFATSVM method, the SVM and TSVM classifiers use a RBF kernel. The Gaussian kernel parameter, $\gamma$, is determined within the range of $[2^{-10}, 2^{-9}, \ldots, 2^9, 2^{10}]$ and the penalty parameter, $C$, is determined within the range of $[2^{-5}, 2^{-4}, \ldots, 2^9, 2^{10}]$ using a grid search. The parameter combination of $\gamma$ and $C$, which has better performance, selected using five-fold cross-validation. We constructed a SOM neural network with $25 \times 25$ neurons arranged in a hexagonal lattice on the output layer (map). The network was trained using all of the available input patterns. The network was trained for 4000 epochs. The initial value of the neighborhood function was set to 20 and gradually decreased, reaching one at the end of 4000 epochs.

To demonstrate the effectiveness of the method proposed, we compare several classical AL methods and SSL methods with this experimental method. The AL methods used for comparison are the MCLU-ECBD, BvSB, entropy, and random methods. For the uncertainty sampling step, $m$ samples are selected, and then $h$ ($1 < h < m$) samples are selected from the $m$ samples in the diversity sampling step. The number of samples for the different AL methods is shown in Table 3.

The AL and semi-supervised combination methods used for comparison are the CASSL and AL + LCR_MDSSSL methods, as well as the BvSB-ECBD-TSVM method.
which combines AL and a TSVM sequentially. The number of AL samples, $q_l$, in the CASSL algorithm is set to 3; in the AL + LCR_MD SSL method, the number of AL samples is set to 1, the number of AL iterations, $M$, is set to 3, and the number of semi-supervised iterations, $K$, is set to 30.

The BvSB-ECBD-TSVM method is divided into two steps. The first step uses AL and the BvSB-ECBD sampling strategy, where the uncertainty sampling parameter, $m$, is set to 12 and the diversity selection parameter, $h$, is set to 3. After the AL part is completed, the TSVM SSL part is performed. The SSL sampling number, $t$, is set to 12, $q_L$ is set to 6 and the AL iteration number, $N/U/M_{AL}$, is set to 8, so the SSL iteration number, $K$, is 22.

Every algorithm runs for 30 iterations on each dataset. To rule out the impact of the randomness of the initial training samples on the experimental results, we repeated 10 experiments for each method. The average accuracy results of the 10 experiments were used as the final precision results to compare the classification performance of the method proposed with the other methods using the same hyperspectral sea ice remote-sensing dataset.

In the experiment, we used a batch algorithm to train the SOM neural network. This algorithm was implemented using MATLAB functions. A multiclass SVM with an OAA architecture was manually implemented using the library for support vector machines, LIBSVM (for the MATLAB interface). The TSVM training and classification process uses the SVMlight library (for the MATLAB interface). All of the algorithms presented in this paper have been implemented in MATLAB.

3.3. Results: Baffin Bay dataset (Experiment 1)

This section contains the classification results and analysis of the Baffin Bay dataset. To demonstrate the advantages of the AL + SSL method over the AL method, we compare the results of the CFATSVM method with the AL methods in Section 3.3.1. Section 3.3.2 shows the results of the CFATSVM method and the other AL + SSL methods, and the classification map of the CFATSVM method for the Baffin Bay dataset is shown in Section 3.3.3.

3.3.1. Comparison between the CFATSVM and AL methods

Figure 4 shows the average overall classification accuracy achieved by the different methods versus the number of samples included in the training set at different iterations for the Baffin Bay dataset. From the figure, it can be seen that, relative to the random method, different AL methods can improve the classification accuracy to various degrees by querying the more informative and diverse samples to label. However, because the sampling strategies of different AL methods are different, the samples selected lead to different levels of classification performance. For example, the BvSB, entropy and MCLU-ECBD algorithms are classical AL algorithms, but their classification results are not very satisfactory, and the final classification accuracy is 73.38%, 76.17% and 89.73%, respectively. As the number of iterations increases, the classification accuracy of the random method fluctuates greatly, and the final precision is relatively low, only 71.57%.

From Fig. 4, compared to the other AL methods (BvSB, entropy and MCLU-ECBD), the CFATSVM method proposed achieves the highest classification accuracy and has a much faster convergence speed. For example, the CFATSVM method achieves a classification accuracy of 90.00%, almost reaching the final convergence accuracy, at the 5th iteration. The MCLU-ECBD algorithm needs 28 iterations to achieve a similar accuracy result. The final classification accuracy of the BvSB and entropy algorithms is also much lower than that of the CFATSVM method over all iterations. This proves that the method proposed can achieve a higher classification accuracy than the other methods using the same labeling samples. At the 11th iteration, the CFATSVM algorithm had an accuracy of 89.99%, the MCLU-ECBD algorithm had an accuracy of 83.90%, and the other three methods had even lower accuracy results. The same accuracy was obtained using substantially fewer labeling samples: For an accuracy of 84%, the CFATSVM algorithm required three more iterations, the MCLU-ECBD algorithm required 13 more iterations, and the other methods required even more iterations.

From the experimental results, it is clear that the method proposed using the CFATSVM algorithm achieves superior classification performance. The algorithm applies an uncertainty sampling strategy based on the BvSB algorithm and a diversity sampling strategy based on the SOM and ECBD methods. Therefore, not only can the most informative sam-
amples be selected, but also the representative cluster samples can be chosen by fully excavating the diversity among the samples. The method proposed selects the most valuable samples with less redundancy, so it is more conducive to improving the accuracy of sea ice classification. On the other hand, the CFATSVM algorithm incorporates the concept of SSL and makes full use of a large number of unlabeled samples. Accordingly, not only is the classification accuracy of the sea ice improved, the labeling cost is significantly reduced. This indicates that these unlabeled samples can better characterize the distribution characteristics of different types of sea ice, and thus improve classification accuracy.

3.3.2. Comparison between the CFATSVM and AL + SSL methods

Figure 5 shows the average classification accuracy results for 10 runs of the testing set using the CFATSVM algorithm proposed and the results using other AL + SSL algorithms. As shown in the figure, the CFATSVM method has obvious advantages over the other methods. The precision of the CFATSVM algorithm converges at the fifth iteration, and the accuracy is very stable as the number of iterations increases. The CASSL method converges relatively slowly, reaching convergence at the tenth iteration, and the accuracy fluctuates slightly in the subsequent iterations. The BvSB-ECBD-TSVM method first implements AL and then a SSL process. Due to the lack of a verification mechanism, this method adds some false label samples, which leads to a lower final classification accuracy. It can also be seen from the figure that the accuracy results obtained using the BvSB-ECBD-TSVM method reach convergence only after many iterations, and the final accuracy results are 1.01% lower than that of the CFATSVM method. The AL + LCR_MD SSL method uses a label changing rate (LCR) to select samples with high reliability for SSL. However, the class center samples selected that have a median distance (MD) to the current SVM classifying sea ice, the experimental results in Fig. 5 verify the correctness of the method.

3.3.3. Sea ice classification results

To visualize the results of the CFATSVM method for classifying sea ice, Fig. 6 shows the results of the sea ice classification experiment on some areas of Baffin Bay in Greenland. Figure 6(a) is an original hyperspectral image composed of bands 159, 194 and 208. Figure 6(b) is the classification image of the Landsat-8 data obtained using a standard SVM classifier, which is used as a reference image for evaluating classification performance. Further, Fig. 6(c) is the classification image of the hyperspectral image obtained using the CFATSVM method. To further evaluate the classification performance of the algorithm in this paper, the overall classification accuracy was calculated for the entire hyperspectral image. The accuracy achieved was 90.003% and the final Kappa coefficient was 0.6966. These results show that the labels predicted are highly consistent with those of labels that were validated.

3.4. Results: Bohai Bay dataset (Experiment 2)

In order to verify the effectiveness of the CFATSVM method, we conducted another experiment using the Bohai Bay dataset. Section 3.4.1 analyzes the results of the CFATSVM method and different AL methods, and Section 3.4.2 shows the results of the CFATSVM method and other AL + SSL methods. The classification map of the uncertainty and diversity strategies, the algorithm not only selects the most informative samples, but also removes a large number of redundant samples, which ensures that the most valuable samples are chosen. In the learning phase between the AL and SSL parts, the consistency verification procedure for the AL and SSL classification results ensures the correctness of the pseudo-labeled samples added. Therefore, compared with other combination methods, the method proposed can achieve rapid convergence and a high classification accuracy, making it suitable for the remote-sensing image classification of sea ice. The experimental results in Fig. 5 verify the correctness of the method.
CFATSVM method when using the Bohai Bay dataset is shown in Section 3.4.3.

3.4.1. Comparison of CFATSVM and AL methods

To further validate the method proposed, we conducted another experiment using the Bohai Bay dataset. Figure 7 shows the overall classification accuracy for the Bohai Bay dataset using different methods versus the number of iterations. From the figure, it can be seen that, relative to the random method, different AL methods can improve the classification accuracy to various degrees by querying more valuable samples for labeling. For example, the BvSB, entropy and MCLU-ECBD algorithms are classic AL algorithms, but their classification results are not very satisfactory, and the final classification accuracy is 94.70%, 94.64% and 97.13%, respectively. With an increasing number of iterations, the classification accuracy of the random method improved slowly, and the final precision was only 92.07%.

The CFATSVM method proposed achieves the highest classification accuracy and has a much faster convergence speed compared to the other AL methods in Fig. 7. For example, the CFATSVM method achieves a classification accuracy of 97.27% at the 13th iteration, and the MCLU-ECBD algorithm needs 30 iterations to achieve a similar accuracy. The classification accuracy results achieved using the BvSB and entropy algorithms are lower than that of the CFATSVM method at every iteration. The CFATSVM algorithm adopts the uncertainty sampling strategy of the BvSB algorithm and the diversity sampling strategy of the SOM and ECBD algorithms. This combination not only queries the most informative samples, but also chooses representative cluster samples by fully excavating the differences among the samples. The method proposed selects the most valuable samples with less redundancy, so it is more conducive to improving the accuracy of sea ice classification. Additionally, the CFATSVM algorithm utilizes a semi-supervised algorithm, and makes full use of a large number of unlabeled samples to improve the classification accuracy of the sea ice and substantially reduce labeling cost. This indicates that the unlabeled samples contain the distribution characteristics of different types of sea ice and thus improve classification accuracy.

3.4.2. Comparison of CFATSVM and AL + SSL methods

From Fig. 8, it can be seen that the CFATSVM method has obvious advantages over other methods when comparing the test set over 10 runs. The accuracy of the CFATSVM algorithm is very stable after the 13th iteration. The CASSL method did not converge until the 27th iteration. The BvSB-ECBD-TSVM method, lacking a verification mechanism, added some false labels to the unlabeled samples, which led to a lower final classification accuracy. It can also be seen from the figure that the accuracy results obtained using the BvSB-ECBD-TSVM method reach convergence only after 21 iterations, and the final accuracy result is 1.21% lower than that of the CFATSVM method.

The AL + LCR_MDS SSL method uses LCR to select samples with high reliability for SSL. However, the class center samples with a MD to the current classification hyperplane contain less information than the samples selected by the AL method used in the CFATSVM algorithm. Additionally, it’s possible to add samples that have incorrect labels, which reduces the performance of the classifier. The final accuracy of 95.38% is lower than that of the CFATSVM algorithm and is consistent with our analysis.

Figure 8 shows that the CFATSVM method proposed achieves good overall performance. While achieving the
same accuracy, the CFATSVM method requires fewer iterations than other methods, which reduces the cost of manual annotation to a certain extent. With the same number of iterations, the CFATSVM algorithm can achieve better precision than the other methods. For example, at five iterations, the CFATSVM algorithm can achieve better precision than other methods. Therefore, compared with other AL + SSL methods, the CFATSVM method requires fewer iterations than other methods, which reduces the cost of manual annotation to a certain extent. With the same number of iterations, which reduces the cost of manual annotation to a certain extent. With the same number of iterations, which reduces the cost of manual annotation to a certain extent.

Fig. 8. Average classification accuracy for the testing dataset over 10 runs using four AL + SSL methods (CFATSVM, CASSL, BvSB-ECBD-TSVM, AL + LCR_MD SSL).

Table 4. The classification accuracy of the CFATSVM method using 126 labeled samples with different values of \( t \) and \( q_2 \) on the Baffin Bay dataset.

| \( t = q_1 \) | \( t = 2q_1 \) | \( t = 3q_2 \) |
|--------------|--------------|--------------|
| \( q_2 = 3 \) | 98.90% | 98.91% | 98.91% |
| \( q_2 = 6 \) | 98.98% | 98.99% | 98.96% |
| \( q_2 = 9 \) | 98.94% | 98.96% | 98.93% |

Table 5. The classification accuracy of the CFATSVM method using 126 labeled samples with different values of \( t \) and \( q_2 \) on the Bohai Bay dataset.

| \( t = q_1 \) | \( t = 2q_1 \) | \( t = 3q_2 \) |
|--------------|--------------|--------------|
| \( q_2 = 3 \) | 97.16% | 97.16% | 97.10% |
| \( q_2 = 6 \) | 97.16% | 97.23% | 97.16% |
| \( q_2 = 9 \) | 96.67% | 97.11% | 96.92% |

Table 4 and Table 5 show the classification accuracy of the CFATSVM method using 126 labeled samples with different values of \( t \) and \( q_2 \) on the Baffin Bay dataset and the Bohai Bay dataset. From the tables, it can be seen that when \( q_2 = 6 \), the final accuracy obtained using the CFATSVM method is 96.675%, which is higher than the accuracies obtained using other methods. The results show that the labels predicted by the CFATSVM method are highly consistent with those of the other methods.

3.4.3. Sea ice classification results

To visualize the results of the CFATSVM method on the classification of sea ice, we randomly select some areas of the Bohai Bay as an example. Figure 9(a) is an original hyperspectral image composed of bands 29, 23, and 16. Figure 9(b) shows the classification image of the Landsat-7 data obtained using a standard SVM classifier, which is used as a reference image for evaluating the classification performance. Figure 9(c) shows the classification image of the hyperspectral image obtained using the CFATSVM method. To further evaluate the classification performance of the algorithm in this paper, the overall classification accuracy was calculated for the entire hyperspectral image, which was 97.24%, and the corresponding Kappa coefficient was 0.9491. The results show that the labels predicted by the CFATSVM method are highly consistent with those of the other methods.

3.5. Parameter sensitivity analysis

The CFATSVM method uses five parameters: \( h_2, h_1, q_1, t \), and \( q_2 \). Among them, \( h_2, h_1, q_1 \) are the parameters used in the sampling strategy, \( Q_1 \). According to Han et al.,\(^{10}\) we set \( q_1 = 3 \) and \( h_2 = 12 \). According to Experiment 5 of Patra and Bruzzone,\(^{19}\) we set \( h_2 = 2h_1 \), that is, \( h_1 = 6 \). In addition, we analyzed the performance of the method proposed by varying the values of \( t \) and \( q_2 \) in the sampling strategy, \( Q_2 \). Table 4 and Table 5 show the classification results obtained using the CFATSVM method considering different values of \( t \) and \( q_2 \) for the Baffin Bay dataset and the Bohai Bay dataset. From the tables, it can be seen that when \( q_2 = 6 \), the final accuracy obtained using the CFATSVM method is 96.675%, which is higher than the accuracies obtained using other methods. The results show that the labels predicted by the CFATSVM method are highly consistent with those of the other methods.
always greater than that of other $q_2$ values. This is because when $q_2$ is very small (i.e., the number of TSVM unlabeled training samples is small), there is not much assistance in improving the performance of the classification model. When $q_2$ is larger, incorrect labels are introduced into the model, resulting in a larger deviation in TSVM parameter $N'$, thus affecting the final classification accuracy of the model. It can also be seen from the tables that when $t = 2q_2$, the accuracy of the CFATSVM algorithm is greater than that when $t$ is some other multiple of $q_2$. This is because when $t = q_2$, the TSVM unlabeled training samples do not undergo the process of diversity selection. Additionally, there are samples with information redundancy, which is not sufficient for improving classification model performance. When $t = 3q_2$, the model will introduce incorrect labels, which affects the final classification accuracy of the model. Therefore, $q_2 = 6$ and $t = 12$ are suitable choices for the sampling strategy, $Q_2$, of the CFATSVM method proposed.

4. Conclusions

Aiming at the problems of high marking costs, difficult to obtain labeled samples and the low accuracy of sea ice classification, this paper proposes a cooperative classification framework based on AL and SSL for sea ice remote-sensing image classification. To evaluate the effectiveness of the method proposed, we conducted experiments on two hyperspectral remote-sensing sea ice image datasets and compared the method proposed with other AL methods and collaborative classification methods. The experimental results show that, starting from a small number of label samples, the CFATSVM method proposed can effectively utilize a large number of unlabeled samples containing abundant information and achieve better classification performance in general. From the experimental results, we summarize as follows.

AL can use fewer unlabeled training samples to automatically select a batch of valuable samples with abundant information to participate in the training sample set and improve the accuracy of the classifier. Different AL methods are based on different sampling strategies, leading to different final classification accuracy results. The approach proposed in this paper combines the AL part and BVSB, SOM and ECBD methods to query representative cluster samples having abundant information and non-redundancy, which are found in the low-density area of large unlabeled samples. Through manual annotation, training sample set labels are included, and the sampling strategy demonstrates effective AL.

SSL can make full use of the distribution information of unlabeled samples, further improving classification accuracy. The SSL part of the method proposed integrates a sampling strategy similar to that of AL and uses the TSVM method to select relatively reliable samples containing certain information from a large number of unlabeled samples for iterative training. This can further reduce labeling cost and improve classification accuracy.

AL and SSL are inherently consistent for reducing the cost of manual labeling and improving classification accuracy.

The CFATSVM classification framework proposed in the paper can effectively use AL to select the most valuable samples. On the other hand, it can also make full use of the information contained in a large number of unlabeled samples for training the classification model. The CFATSVM method then verifies the pseudo-labeled samples added to the training set through the collaborative mechanism of AL and SSL, and further improves classification accuracy. The method achieves good classification performance for sea ice classification while lowering the cost of manual labeling and providing a new method for classifying sea ice images.

References

1) Xi, Z., Jie, Z., and Junmin, M.: Polarimetric Scattering Characteristics Based Sea Ice Types Classification by Polarimetric Synthetic Aperture Radar: Taking Sea Ice in the Bohai Sea for Example, Acta Oceano-litica Sinica, 35, 5 (2013), pp. 95–101.
2) Laine, V.: Antarctic Ice Sheet and Sea Ice Regional Albedo and Temperature Change, 1981–2000, from AVHRR Polar Pathfinder Data, Remote Sens. Environ., 112, 3 (2008), pp. 646–667.
3) Kwok, R.: Identification of Sea Ice Types in Spaceborne Synthetic Aperture Radar Data, J. Geophys. Res., 97, C2 (1992), pp. 2391–2402.
4) Zakhvatkina, N. Y.: Classification of Sea Ice Types in ENVISAT Synthetic Aperture Radar Images, IEEE Trans. Geosci. Remote Sens., 51, 5 (2013), pp. 2587–2600.
5) Dabboor, M. and Geldsetzer, T.: Towards Sea Ice Classification Using Simulated RADARSAT Constellation Mission Compact Polarimetric SAR Imagery, Remote Sens. Environ., 140 (2014), pp. 189–195.
6) Tao, Z.: Classification of Sea Ice Based on Fuzzy Clustering with Spatial Information, Int. Conf. on Multimedia Technology (ICMT), IEEE, 2010, pp. 1–4.
7) Istromana, L. and Heygster, G.: Clouds Discrimination and Surface Classification for the Sea Ice Albedo Retrieval from MODIS Data, IEEE Int. Geoscience and Remote Sensing Symp. (IGARSS), 2012, pp. 4481–4484.
8) Rosel, A. and Kaleschke, L.: Comparison of Different Retrieval Techniques for Melt Ponds on Arctic Sea Ice from Landsat and MODIS Satellite Data, Ann. Glaciol., 52, 57 (2011), pp. 185–191.
9) Cavalieri, D. J.: Assessment of EOS Aqua AMSR-E Arctic Sea Ice Concentrations Using Landsat-7 and Airborne Microwave Imagery, IEEE Trans. Geosci. Remote Sens., 44, 11 (2006), pp. 3057–3069.
10) Han, Y. L., Ren, J., and Hong, Z. H.: Active Learning Algorithms for the Classification of Hyperspectral Sea Ice Images, Mathematical Problems in Engineering, Vol. 2015. Accessed on: 2015, DOI:10.1155/2015/124601 [Online].
11) Settles, B.: Active Learning Literature Survey, University of Wisconsin-Madison, 39, 2 (2009), pp. 127–131.
12) Zhu, X. J.: Semi-supervised Learning Literature Survey, University of Wisconsin-Madison, 37, 1 (2005), pp. 63–77.
13) Liu, R., Wang, Y., and Baba, T.: SVM-Based Active Feedback in Image Retrieval Using Clustering and Unlabeled Data, Pattern Recognit., 41, 8 (2008), pp. 2645–2655.
14) Joshi, A. J., Porikli, F., and Papanikolopoulos, N.: Multi-class Active Learning for Image Classification, 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, 2009, pp. 2372–2379.
15) Denm, B., Persello, C., and Bruzzone, L.: Batch-Mode Active-Learning Methods for the Interactive Classification of Remote Sensing Images, IEEE Trans. Geosci. Remote Sens., 49, 3 (2011), pp. 1014–1031.
16) Chen, R., Cao, Y. F., and Sun, H.: Multi-class Image Classification with Active Learning and Semi-supervised Learning, Acta Automatica Sinica, 37, 8 (2011), pp. 954–962.
17) Patra, S. and Bruzzone, L.: A Fast Cluster-Association Based Active-Learning Technique for Classification of Remote Sensing Images, IEEE Trans. Geosci. Remote Sens., 49, 5 (2011), pp. 1617–1626.
18) Copa, L., Tuia, D., and Volpi, M.: Unbiased Query-by-bagging Active
Learning for VHRI Image Classification, Image and Signal Processing for Remote Sensing XVI, International Society for Optics and Photonics, Vol. 7830, 2010, pp. 176–183.

19) Patra, S. and Bruzzone, L.: A Novel SOM-SVM-Based Active Learning Technique for Remote Sensing Image Classification, *IEEE Trans. Geosci. Remote Sens.*, **52**, 11 (2014), pp. 6899–6910.

20) Shi, Q., Liu, X., and Huang, X.: An Active Relearning Framework for Remote Sensing Image Classification, *IEEE Trans. Geosci. Remote Sens.*, **56**, 6 (2018), pp. 3468–3486.

21) Dópido, I., Li, J., and Marpu, P. R.: Semi-supervised Self-Learning for Hyperspectral Image Classification, *IEEE Trans. Geosci. Remote Sens.*, **51**, 7 (2013), pp. 4032–4044.

22) Camps-Valls, G., Marshev, T. V. B., and Zhou, D.: Semi-Supervised Graph-Based Hyperspectral Image Classification, *IEEE Trans. Geosci. Remote Sens.*, **45**, 10 (2007), pp. 3044–3054.

23) Zhou, D., Bousquet, O., Lal, T., Weston, J., and Oltropf, B.: Learning with Local and Global Consistency, Advances in Neural Information Processing Systems, Vol.16, 2004.

24) Vapnik, V. and Sternin, A.: On Structural Risk Minimization or Overall Risk in a Problem of Pattern Recognition, *Automation Remote Control*, **10**, 10 (1977), pp. 1495–1503.

25) Chen, Y., Wang, G., and Dong, S.: Learning with Progressive Transductive Support Vector Machine, *Pattern Recognition Letters*, **24**, 10 (2003), pp. 1845–1855.

26) Zhou, Z. H. and Li, M.: Semi-supervised Learning by Disagreement, *Knowledge Inform. Syst.*, **24**, 3 (2010), pp. 415–439.

27) Bruzzone, L., Cha, M., and Marconcini, M.: A Novel Transductive SVM for Semi-supervised Classification of Remote-Sensing Images, *IEEE Trans. Geosci. Remote Sens.*, **44**, 11 (2006), pp. 3363–3373.

28) Aydav, P. S. S. and Minz, S.: Classification of Hyperspectral Images Using Self-training and a Pseudo Validation Set, *Remote Sens. Lett.*, **9**, 11 (2018), pp. 1109–1117.

29) Aydav, P. S. S. and Minz, S.: Granulation-based Self-training for the Semi-supervised Classification of Remote-sensing Images, *Granular Computing*, 2019, doi:10.1007/s41066-019-00161-x

30) Zhang, Z. and Crawford, M. M.: A Batch-Mode Regularized Multimetric Active Learning Framework for Classification of Hyperspectral Images, *IEEE Trans. Geosci. Remote Sens.*, **55**, 11 (2017), pp. 6594–6609.

31) Wan, L., Tang, K., Li, M., Zhong, Y., and Qin, A. K.: Collaborative Active and Semi-supervised Learning for Hyperspectral Remote Sensing Image Classification, *IEEE Trans. Geosci. Remote Sens.*, **53**, 5 (2015), pp. 2384–2396.

32) Leng, Y., Xu, X., and Qi, G.: Combining Active Learning and Semi-supervised Learning to Construct SVM Classifier, *Knowledge-Based Syst.*, **44**, 1 (2013), pp. 121–131.

33) Zhou, X. and Prasad, S.: Active and Semi-supervised Learning with Morphological Component Analysis for Hyperspectral Image Classification, *IEEE Geosci. Remote Sens. Lett.*, **99** (2017), pp. 1–5.

34) Liu, C., Li, J., and He, L.: Superpixel-Based Semi-supervised Active Learning for Hyperspectral Image Classification, *IEEE J. Selected Topics Appl. Earth Observ. Remote Sens.*, **12**, 1 (2019), pp. 357–370.

35) Persello, C. and Bruzzone, L.: Active and Semisupervised Learning for the Classification of Remote Sensing Images, *IEEE Trans. Geosci. Remote Sens.*, **52**, 11 (2014), pp. 6937–6956.

36) Barry, P., Segal, C., and Carman, S.: EO-1/Hyperspectral Science Data User’s Guide, 2001.

37) Du, Q. and Yang, H.: Similarity-Based Unsupervised Band Selection for Hyperspectral Image Analysis, *IEEE Geosci. Remote Sens. Lett.*, **5**, 4 (2008), pp. 564–568.

38) Han, Y., Li, J., and Zhang, Y.: Sea Ice Detection Based on an Improved Similarity Measurement Method Using Hyperspectral Data, *Sensors*, **17**, 5 (2017), p. 1124.

39) Hossain, M. S., Bujang, J. S., and Zakaria, M. H.: Assessment of Landsat 7 Scan Line Corrector-off Data Gap-filling Methods for Seagrass Distribution Mapping, *Int. J. Remote Sens.*, **36**, 4 (2015), pp. 1188–1215.

40) Han, Y. L., Li, P., and Zhang, Y.: Combining Active Learning and Transductive Support Vector Machines for Sea Ice Detection, *J. Appl. Remote Sens.*, **12**, 2 (2018), p. 1.

41) Chang, C.-C. and Lin, C.-J.: LIBSVM: A Library for Support Vector Machine, 2001 [Online]. Available: http://csie.ntu.edu.tw/~cjlin/libsvm

42) Joachims, T.: SVMlight: Support Vector Machine, 2008, http://www.cs.cornell.edu/people/tj/svm.light

Zheng Hong Zhu

Associate Editor