Received July 9, 2019, accepted July 26, 2019, date of publication August 6, 2019, date of current version September 16, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2933589

A New Framework for Hyperspectral Image Classification Using Multiple Semisupervised Collaborative Classification Algorithm

YING CUI, XIAOWEI JI, (Student Member, IEEE), HENG WANG, KAI XU, SHAOQIAO WU, AND LIGUO WANG, (Member, IEEE)

College of Information and Communications Engineering, Harbin Engineering University, Harbin 150001, China

Corresponding author: Xiaowei Ji (jixiaowei@ hrbeu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61675051 and fundamental scientific research funds for the Central Universities under Grant 3072019CF0801.

ABSTRACT Hyperspectral images (HSIs) have evident advantages in image understanding because of enormous spectral bands, and rich spatial information. However, applying the limited labeled samples to obtain satisfactory classification results is a challenging task. Secondary screening algorithm and semisupervised learning are two promising methods to address this problem. Secondary screening algorithm exploits different query functions, which are on the basis of the evaluation of two criteria: uncertainty and diversity. The advantage of semisupervised learning is that with a small number of samples, classifiers could learn the structure of whole data sets without significant costs and efforts. Hence, combining secondary screening algorithm and semisupervised learning is a natural consideration. We firstly investigate nine secondary screening algorithms and compare their performance. Next, two novel frameworks are proposed in this paper. They are named the syncretic one-fold secondary screening algorithm and semisupervised learning framework (OFSS-SL) and syncretic multiple secondary screening algorithms and multiple-verification semisupervised learning framework (MSS-MVSL), respectively. We evaluate the performance of OFSS-SL and MSS-MVSL on three hyperspectral data sets and compare them with that of three state-of-the-art classification methods. In general, our results suggest that two proposed frameworks can apply limited labeled samples to achieve excellent classification results. And the computational costs of them are cheaper than previous methods.

INDEX TERMS Active learning, hyperspectral image classification, semisupervised learning.

I. INTRODUCTION

With the development of spectral imaging techniques and optical sensor systems in recent years, hyperspectral images, which contain a lot of information in hundreds of continuous and narrow spectral bands, have been widely used for many applications, such as image fusion, change detection, and classification, etc [1]–[3]. To better characterize complex scenes of remote sensing images, classification is a crucial processing step in many applications. It is widely acknowledged that a hyperspectral image data set usually contains limited labeled samples and a large number of unlabeled samples. As for supervised learning, the quality and quantity of the labeled samples determine the performance of the classification. A variety of machine learning techniques have been used for HSI classification [4], including support vector machines [5], [6], neural networks [7], [8] and regression methods [9], [10]. However, labeling a large number of samples is time-consuming and obtaining high-quality labeled samples is a difficult task. Therefore, a direct idea is designing a technique that requires little or no labeled training data. On the other end of the spectrum from classification, unsupervised learning also has ability to address spectral-spatial joint feature calculation and classification. Whereas, the problem of unsupervised learning of hyperspectral images is also a significant challenge [11]. Clustering technique is usually applied in unsupervised learning for data classification. Recent techniques for hyperspectral clustering include these based on particle swarm optimization [12], density analysis [13], nearest neighbor clustering [14], total variation methods [15], and sparse manifold models [16], [17].

The associate editor coordinating the review of this article and approving it for publication was Shagufta Henna.

How to use the least labeled samples to minimize the time consumption and improve the performance of classifiers have become crucial issues in hyperspectral image classification. Active learning (AL), semisupervised learning (SSL) and clustering techniques provide promising methods to improve the performance. Unlike supervised learning, semisupervised learning can effectively enhance the classification result by deploying limited labeled samples [18], [19]. This can be attributed to the fact that semisupervised learning pays more attention to the unlabeled data in unsupervised approach. Simultaneously, semisupervised learning facilitates the supervised model by increasing the quantity of the training samples, and improves the generalization ability of the classifier by applying both labeled data and unlabeled data. Another interesting modality is active learning. As a subfield of machine learning, AL has been successfully applied to select training sample sets. Because it can minimize the number of training samples and keep the discriminative capabilities of classifiers as high as possibly [20], [21]. The AL process is conducted according to an iterative process. It utilizes a query function to inspect the whole unlabeled data and select the most informative one or a batch of samples for manual labeling at each iteration, then the labeled data set and the unlabeled data set are updated by adding the new labeled samples to the labeled set and removing them from the unlabeled set. AL can obtain high quality classification results with fewer labeled samples than in the case of randomly selected training samples.

On the basis of the aforementioned considerations, in order to mitigate the computational costs and improve the pseudolabeling accuracy, we investigate different AL techniques to build a buffer pool which embrace the batch of the most informative sample, then apply the different clustering algorithms for secondary selection, which confirm the selected unlabeled are diverse. Moreover, we propose two novel semisupervised learning frameworks which combine unsupervised secondary screening algorithms for guaranteeing the diversity of samples. The first one is syncretic one-fold secondary screening algorithm and semisupervised learning framework (OFSS-SL). The second one is syncretic multiple secondary screening algorithms and multiple-verification semisupervised learning framework (MSS-MVSL). The investigated techniques and proposed frameworks are compared with related algorithms in the classification of hyperspectral images.

The results on three widely used hyperspectral data sets verify the effectiveness of the proposed frameworks. Compare with the classical collaborative active and semisupervised learning method (CASSL), the proposed frameworks can effectively train classifiers and sufficiently improve the classification performance with small sized number of labeled samples. Different from single batch-mode active-learning methods, such as MCLU-ECBD, we implement multiple secondary screening strategy to mine representative and discriminative information from the unlabeled samples, the most valuable samples also are diverse. Simultaneously, by multiple validation part, we can acquire more confidently pseudolabeled samples to facilitate SSL.

The rest of this paper is organized as follows. Section II presents details of the related works on hyperspectral image classification. In Section III, first, we specifically introduce the methods applied in this paper. Then we exploit and compare nine unsupervised secondary screening algorithms, which mine the representative information and diversity of samples. In Section IV, we propose two novel frameworks. Section V specifically describes the effectiveness of two proposed frameworks from experimental results. Section VI summarizes this paper.

II. RELATED WORK

The aim of this section is to briefly introduce the present active learning algorithms and clustering techniques developed for hyperspectral image classification. It is, moreover, important to technically precise the relationships between semisupervised learning and active learning and exploit the intimately related watershed.

A. ACTIVE LEARNING FOR HYPERSPECTRAL IMAGE CLASSIFICATION

Active learning is an effective method to solve the problem of less training samples, which is a challenge for hyperspectral image classification. According to the active learning literatures, an active learner can be modeled as a quintuple [22], Q is a query function and applied to select the most informative unlabeled samples from unlabeled data set and the unlabeled data set is denoted as U. At the initial stage, an initial training set T of few labeled samples is required for the first training of the classifier G. S is a supervisor who can assign the true class label to the unlabeled samples.

At the beginning of the iteration loop, active learning utilizes a query function to select the most informative samples, then these most informative samples are labeled by human experts. At the same time, the labeled data set and the unlabeled data set are updated. In this way, only the most informative samples are to be labeled, the noninformative samples are ignored. Hence, the time consumption and the cost of data collection can be significantly decreased. From this procedure, we observe that the query function is a key to the active learning. There are two main criteria used to design a query function: the uncertainty and the diversity [23]–[26]. Recently, there are many researches on the uncertainty criterion, such as query by committee, the posterior probability-based methods and the large margin heuristic-based methods. Query by committee which measures the uncertainty of the samples by the maximum variance among the committee of learners [27]–[29]. Entropy provides a method to study and understand the variation of uncertainty in classification outputs. Tuia *et al.* [30] proposed an active learning algorithm with the help of the entropy query-by-bagging algorithm (EQB), and it is independent on the classifiers and performs well on remote sensing data sets. EQB has the tendency to locally over sample complex areas, leading to search the regional feature space. Li *et al.* [31] proposed an improved

EQB algorithm named averaged entropy query-by-bagging algorithm (aEQB). Copa *et al.* [32] proposed a normalized entropy query-by-bagging algorithm (nEQB). This method yields more reliable results than EQB. Compared with the query by committee, Rajan *et al.* [33] firstly proposed an active-learning method based on posterior probability. This method by contrasting the maximum difference of the posterior probability between before and after adding the samples to the training set. The information gain is measured by the Kullback–Leibler (KL) divergence. The KL-Maximization (KL-Max) technique can be implemented with any classifier, but it only select one sample at each iteration. As for the large margin heuristic-based methods, Mitra *et al.* [34] proposed an active SVM learning method based on margin sampling (MS) for object-oriented remote sensing image segmentation. This method selects the data points that are closest to the current separating hyperplane. In order to solve the problem of multi-classification, Multiclass level uncertainty (MCLU) was be proposed by Demir *et al.* [35] and Shi *et al.* [36]. This technique effectively selects the most informative samples according to the confidence values $c(x)$ and is widely used in remote-sensing image classification.

B. CLUSTERING TECHNIQUES FOR HYPERSPECTRAL IMAGE CLASSIFICATION

Clustering is also an essential step in many ways, such as data mining and pattern recognition, which aims to divide a large number of unlabeled data into several clusters based on similarity [11], [38], [39]. By evaluating the distribution of the samples in a feature space, clustering techniques assign the same label for the data in the same cluster, and it is an effective technique to exploit representative information in large-scale data set. Many clustering methods have been presented in the literatures. K-means clustering is a commonly used data clustering for performing unsupervised learning tasks [40]. It should be underlined that k-means clustering offers no accuracy guarantees, but the simplicity and efficient of k-means are very appealing in practice. Fuzzy c-means (FCM) [41] is one of the most widely used methods in fuzzy clustering. Given a certain cluster number, it can find the hidden structure of a data set through optimization of the objective function. Hierarchical clustering constructs a hierarchy of clusters by either iteratively integrating two smaller clusters into a larger one or dividing a larger cluster into smaller ones. And the crucial step of Hierarchical clustering is how to best select the next cluster to split or merge [42]. Compared with other clustering techniques, spectral clustering has good performance in dealing with irregularly-shaped clusters and gradual variation within groups. Because its capability of high-quality clustering and handling non-convex clusters that are typically be superior to many methods [43]. This method takes the similarity matrix as the input and applies k-means to top eigenvectors of the graph Laplacian matrix in the clustering process. Nowadays, spectral clustering has been utilized in many domains such as computer vision, classification, and speech separation with promising performance [44]–[46].

Compared with the uncertainty, less attention has been paid on the diversity criterion in remote sensing image classification. The main idea of using diversity criterion is to select a batch of samples, which are the most informative and, simultaneously, are diverse from each other. Demir *et al.* [35] first selected m most uncertain samples with several SVM-based techniques and clustered the m samples into h clusters ($h < m$). Then, these selected samples are labeled by human experts. The advantage of this approach is that selecting the most uncertain sample from each cluster greatly ensures the diversity criterion. Patra and Bruzzone [37] presented an algorithm based on SVM and a self-organizing map (SOM) neural network, which firstly selected m most uncertain samples by SVM technique, and then selected h samples that correspond with the SOM mapping neurons from the m samples. Multiclass level uncertainty-enhanced cluster-based diversity method is proposed in [35]. It has been employed in an interactive domain adaptation approach for applying the classifier trained on a remote sensing image to a different but related target image.

C. SEMISUPERVISED LEARNING FOR HYPERSPECTRAL IMAGE CLASSIFICATION

In hyperspectral image classification, during the past few decades, the semisupervised learning methods has drawn great interests in remote sensing [47], [48]. Generally speaking, semi-supervised learning can be classified into the generative model [49], the co-training model [50], the graph-based method [51], [52], etc. Wang *et al.* proposed a novel graph-based semisupervised learning approach based on a linear neighborhood model to propagate the labels from the labeled samples to the whole data set using linear neighborhoods with sufficient smoothness [53]. Bruzzone *et al.* [54] proposed progressive semisupervised SVM. This approach improves the SVM classifiers iteratively by assigning pseudolabels to the unlabeled samples that are closest to the margin bound. In [55], de Morsier *et al.* presented a graph representation that is discriminative of the cluster structure of the data and it assumes multiple possible intersecting manifold. In order to avoid assigning incorrect pseudolabels to unlabeled data by human labeling, in [56], Wang *et al.* proposed a method to address the special problematic characteristics of hyperspectral images. First, as a semisupervised approach, it exploits the structure of unlabeled samples by evaluating the confidence probability of the predicted labels. Then, this method jointly optimizes the classifier parameters and the dictionary atoms by a task-driven formulation, which aims at learning features that are optimal for the trained classifier. Finally, it incorporates spatial information through adding a Laplacian smoothness regularization to the output of the classifier.

D. COMBINATION OF AL AND SSL FOR HYPERSPECTRAL IMAGE CLASSIFICATION

The problem of less training samples has become an active investigation orientation in hyperspectral image

classification. To complement active learning and semisupervised learning for each other becomes a natural choice. Most of the previous studies have demonstrated that the combination of them is an excellent approach for hyperspectral images classification. Dópidoa *et al.* [57] proposed a novel semisupervised active-learning method for urban hyperspectral image classification. Initially, utilizing active learning methods to select the most informative samples with achieving a great improvement in the classification results. Then, the classifier estimates the labels of the selected samples with no extra cost for labeling the selected samples. A novel algorithm embedded into the multi-view active learning was proposed by Di and Crawford [58]. In this paper, the first regularizer studies the intrinsic multi-view information embedded in the hyperspectral data. The second regularizer is based on the ''consistency assumption''. In [59], this method combines AL and SSL to invoke a collaborative labeling process by both human experts and classifiers. Firstly, an AL-based pseudolabel verification procedure is performed for constantly increasing the pseudolabeling accuracy to facilitate SSL. Then, only those unlabeled data with low pseudolabeling confidence will become the query candidates in AL.

It is important to observe that the aforementioned methods have many similarities. Firstly, these methods apply active-learning approach to label informative samples. Then the remaining unlabeled samples are assigned by the current classifiers. In the process of labeling the samples, the performance of the classifiers is improved gradually. However, some other limitations can compromise their effectiveness: 1) They pay more attention on discriminative information and ignored the representative information, leading to the information bias. 2) These methods apply the single classifier as validation model and use single active learning algorithm to select informative samples, which have the disadvantage that the difference between the base classifier and the verification classifier is confined to their updated samples. It should be mentioned that the difference of models can directly decide the reliability of the pseudolabels. It indicates that if the characteristics of base classifier and the verification classifier are similar, which can lead to the algorithm has a limited ability to improve the discriminative information. Meanwhile, these algorithm may be immediately end and cannot reach the desired accuracy.

To alleviate these aforementioned issues, nine unsupervised secondary screening algorithms are exploited, which mine the representative information of samples. Section III specifically introduces the unsupervised secondary screening algorithms.

III. INVESTIGATED UNSUPERVISED SECONDARY SCREENING ALGORITHMS

This section mainly introduces the main structure of the proposed algorithm. Representative information and discriminative information, which denote the quality of the labeled samples, are vital to improve the generalization ability for classifiers. However, it is widely acknowledged that the unlabeled samples are numerous. Therefore, using the unlabeled samples to enhance the training samples is an advisable selection. In order to adequately exploit the representative information, we combine active learning methods and unsupervised clustering techniques. We implement the averaged entropy query-by-bagging algorithm (aEQB), normalized entropy query-by-bagging algorithm (nEQB) and Multiclass level uncertainty algorithm (MCLU), to select informative samples and put these samples into first-level buffer pool. Then, samples in the pool are clustered by clustering methods, in each cluster, we select the most informative samples to label by human experts. As for clustering methods, we apply three typical methods, k-means clustering, spectral clustering algorithms and hierarchical clustering algorithms. Simultaneously, the labeled data set and the unlabeled data set are updated by adding these informative samples to the labeled data set and removing them from the unlabeled data set. We combine three active learning algorithms with three clustering algorithms, respectively. Hence, we obtain nine unsupervised secondary screening algorithms to compare their performance, then sort out the unsupervised secondary screening algorithms with greatest performance.

A. QUERY FUNCTION FOR ACTIVE LEARNING

Entropy is a criterion that summarizes the classification uncertainty in a single number, per pixel, per class or per image. Query by bagging (EQB), proposed by Tuia, is a classifier-independent approach based on the selection of unlabeled samples according to the maximum disagreement between a committee of classifiers. In order to ensure the diversity of committee member, the internal committee classifier adopted Bootstrap sampling [60]. The process of EQB can be divided into the following steps: firstly, the original training set is divided into the some training subsets, the number of training subsets is K. Each training set is implemented to train the one-against-all (OAA) SVM architecture to predict the different labels for each unlabeled sample [34]. After training, for each sample $x_i \in U$ has K labels, which can be seen as a classification frequency and as the probability for the candidate x_i to be labeled in the class ω. We could observe that the entropy of the distribution of the different labels associated to each sample is calculated to evaluate the disagreement among the classifiers. The EQB function is expressed as:

$$\mathrm{x}^{EQB} = \arg\max_{x_i \in U} H\left(x_i\right) \tag{1}$$

$$H\left(x_i\right) = \sum_{\omega=1}^{N_i} p\left(y_i^* = \omega | x_i\right) \log\left[p\left(y_i^* = \omega | x_i\right)\right] \tag{2}$$

y_i^* is the class label predicted for the i-th pixel. N_i is the number of classes predicted for x_i by the committee and $1 \leq N_i \leq N$, where N is the total number of classes. $p\left(y_i^* = \omega | x_i\right)$ is the observed probability to obtain class ω predicted for the candidate. The experiment by Demir *et al.* [35] indicates that,

in batch sampling process, the performance of EQB algorithm has poor performance when the number of samples is small. However, the performance of EQB algorithm is improved when extracts more samples, which means that EQB algorithm only determines the approximate segmentation range of information, and it can't accurately arrange the samples according to the amount of information entropy. For the multi-classification, a sample may has high confidence when the committee classifiers maintains the same prediction. It is also indicate that this sample has low uncertainty and small entropy value. Reversely, a sample may has high uncertainty when the committee classifiers maintains the different prediction. It is assumed that the entropy value of this sample is large. Accordingly, the unlabeled sample with the maximum disagreement be worth labeling by human experts. However, some scholars [61] have found that, with regard to the multi-classification problem, in some cases, the entropy value can't represent the uncertainty of samples. Sometimes, the uncertainty of samples with smaller entropy value will be higher than these of larger entropy value. Han *et al.* [62] proved that information entropy had multi-value bias in theory, due to the fact that in the process of classifier training iterative prediction. EQB has the tendency to locally over sample complex areas, implying a large number of classes among which the committee is uncertain. In classification field, diversity in the training set is a key element to obtain good predictions. Successful studies have been carried out to solve this problem. Copa *et al.* proposed nEQB algorithm in [32]. nEQB adopts a normalization to lead to the identification of more relevant samples independently from the number of classes predicted for each candidate x_i, as shown in (3).

$$\mathrm{x}^{nEQB} = \arg\max_{x_i \in U} \left\{ \frac{H(x_i)}{\log(N_i)} \right\} \tag{3}$$

Li *et al.* [31] put forward an improved algorithm called averaged entropy query-by-bagging algorithm (aEQB). This algorithm can also ensure the diversity of sampling process by averaging the entropy and punish the multi-value attribute by adding classification information. The aEQB as follows:

$$\mathrm{x}^{aEQB} = \arg\max_{x_i \in U} \left\{ \frac{H(x_i)}{N_i} \right\} \tag{4}$$

The essence of nEQB and aEQB are similar, both of them can punish the information entropy from the prediction category perspective, and it is difficult to set the penalty items. In different iteration stages, the performance of the penalty relies on the ability of classifier.

The MCLU technique chooses the most uncertain samples according to a confidence value $c(x)$, $x \in U$, which are defined on their functional distance $\{f_1(x), f_2(x), \ldots\ldots, f_n(x)\}$ to the n decision boundaries of the binary SVM classifiers. The classifiers are built in the one-against-all (OAA) architecture. Then, the confidence value $c(x)$ can be calculated by applying different strategies. Two strategies are commonly used, the first one is the minimum distance

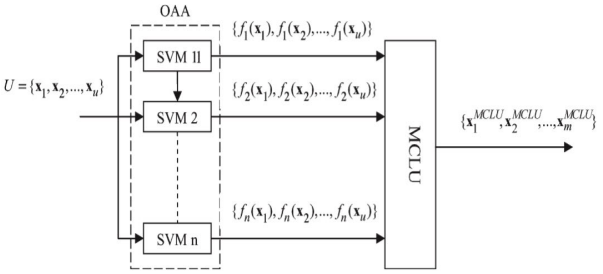

FIGURE 1. Architecture adopted for the MCLU technique.

function $c_{\min}(x)$ strategy. The $c_{\min}(x)$ function applies a simple strategy that computes the confidence of a sample x by calculating the minimum distance to the hyperplanes evaluated on the basis of the most uncertain binary SVM classifier.

$$c_{\min}(x) = \min_{i=1,2,\ldots\ldots n} \{abs[f_i(x)]\} \tag{5}$$

The second one is the difference function $c_{diff}(x)$ strategy, which considers the difference between the first and second largest distance values to the hyperplanes, i.e., [35].

$$r_{1\max} = \operatorname*{argmax}_{i=1,2,\ldots,n} \{f_i(x)\} \tag{6}$$

$$r_{2\max} = \operatorname*{arg\,max}_{j=1,2,\ldots,n, j \neq r_{1\max}} \{f_i(x)\} \tag{7}$$

$$c_{diff}(x) = f_{r1\max}(x) - f_{r2\max}(x) \tag{8}$$

Compared with $c_{\min}(x)$, the $c_{diff}(x)$ strategy compares the uncertainty between the two most likely categories. If the value of $c_{diff}(x)$ is high, the sample x is assigned to $r_{1\max}$ with high confidence. On the contrary, if this value is low, it means that the decision for $r_{1\max}$ is not reliable, and there is a possible contradiction with the class $r_{2\max}$. Hence, this sample is selected by the query function for more appropriate modeling the decision function in the feature space. When the value of $c(x), (x \in U)$ is acquired by one of the two aforementioned strategies, the m samples $x_1^{MCLU}, x_2^{MCLU}, \ldots\ldots, x_m^{MCLU}$ with lower $c(x)$ are selected to be forwarded to the diversity step. It should be mentioned that x_j^{MCLU} represents that the selected j th most uncertain sample based on the MCLU strategy. Fig. 1 shows the architecture of the MCLU technique [35].

Multiclass level uncertainty-enhanced cluster-based diversity (MCLU-ECBD), is employed as the query heuristic. In MCLU-ECBD, the aforementioned MS heuristic is extended to the multiclass scenario using the one-against-all architecture. Specifically, MCLU-ECBD consists of two steps, i.e., MCLU and ECBD. MCLU has been introduced in this section. The ECBD technique works in the kernel space by utilizing the kernel k-means clustering to select the $h < m$ most diverse patterns [64].

In detail, the kernel k-means clustering iteratively divides the m samples into k clusters in the kernel space, $(k = h)$. When the clusters are constructed, assigning initial cluster labels to each sample. In next iterations, a pseudo-center is

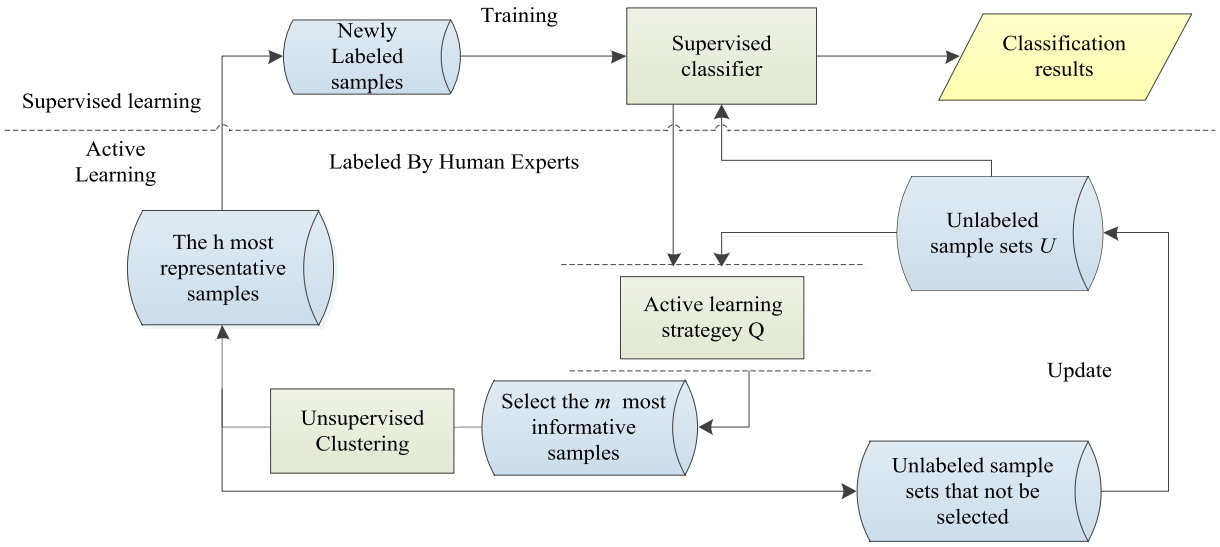

FIGURE 2. Flowchart of the proposed secondary screening algorithms.

chosen as the cluster center. The distance of each sample from all cluster centers in the kernel space is calculated and each sample is assigned to the nearest cluster.

B. DETAILS OF INVESTIGATED METHODS

The flowchart of the investigated algorithm is shown in Fig. 2. It can be divided into two parts, one is the supervised learning of the samples, the other is the act labeling and updating of the unlabeled samples. Moreover, the samples are divided into training set and test set. The test set is used to verify the training effect of the model, and it does not participate in training classifiers. And the training set is divided into labeled set and unlabeled set. We apply the aEQB, nEQB and MCLU, which are widely used to select the most informative samples in remote-sensing image classification. By implementing above-mentioned active learning approaches, we select the top M samples, and puts them into the first level buffer pool. It should be mentioned that, the number of top M samples is m, ($m \geq M$), and we don't label the samples in the first level buffer pool. Then we utilizing different unsupervised clustering techniques to classify them according to the distribution of spectral features of the HSIs. We apply k-means (KM), spectral clustering (SC) and hierarchical clustering (HC), to compresses candidate sample pool. we add the KM, SC and HC to represent the investigated algorithm, respectively. Hence, we evaluate nine possible unsupervised secondary screening algorithms in this paper: 1) MCLU with K-means (denoted by MCLU-KM); 2) MCLU with Spectral clustering algorithms (denoted by MCLU-SC); 3) MCLU with Hierarchical clustering algorithms (denoted by MCLU-HC); 4) nEQB with K-means algorithms (denoted by nEQB-KM); 5) nEQB with Spectral clustering algorithms (denoted by nEQB-SC); 6) nEQB with Hierarchical clustering algorithms (denoted by nEQB-HC); 7) aEQB with K-means algorithms (denoted by aEQB-KM); 8) aEQB with with Spectral clustering algorithms (denoted by aEQB-SC); 9) aEQB with Hierarchical clustering algorithms (denoted by aEQB-HC).

It should be underlined that, in the k-means algorithm, the value of k will determines the effect of clustering, and it is difficult to make a reasonable setting for k value. In this selecting process, the k value sets to h. (h is the number of samples to be labeled). Using clustering algorithm compresses candidate sample pool, to force samples to cluster, and similar samples can be clustered into the same class. The m most informative samples divides into h clusters. Adopting the principle of "first come, first served" to cluster samples from the top-ranking samples. As shown in Fig. 3, for samples selection process, if Data 1 and Data 2 belong to the same class, we only choose one with higher information entropy in the same class. We choose Data 1 and Data 3, although the information of Data 2 is higher than that of Data 3. And we summarize the algorithm in Algorithm 1.

IV. PROPOSED METHODOLOGY

In this section, we propose two novel frameworks that integrate secondary screening algorithms and the semisupervised learning in a collaborative manner for hyperspectral image classification. They are OFSS-SL and MSS-MVSL. It should be stressed that integrating secondary screening algorithms and semisupervised learning is a novel concept and not a simple combination of two methods. The first novelty of proposed framework is that applying the combination of secondary screening algorithms to mine the representative information of unlabeled samples at each iteration. The second novelty is that implementing double verification classifier to learn the composition and distribution of unlabeled data set without significant cost. Then, the classifier approximates the model and structure of the remaining unlabeled data. In this case, proposed framework can avoid introducing many

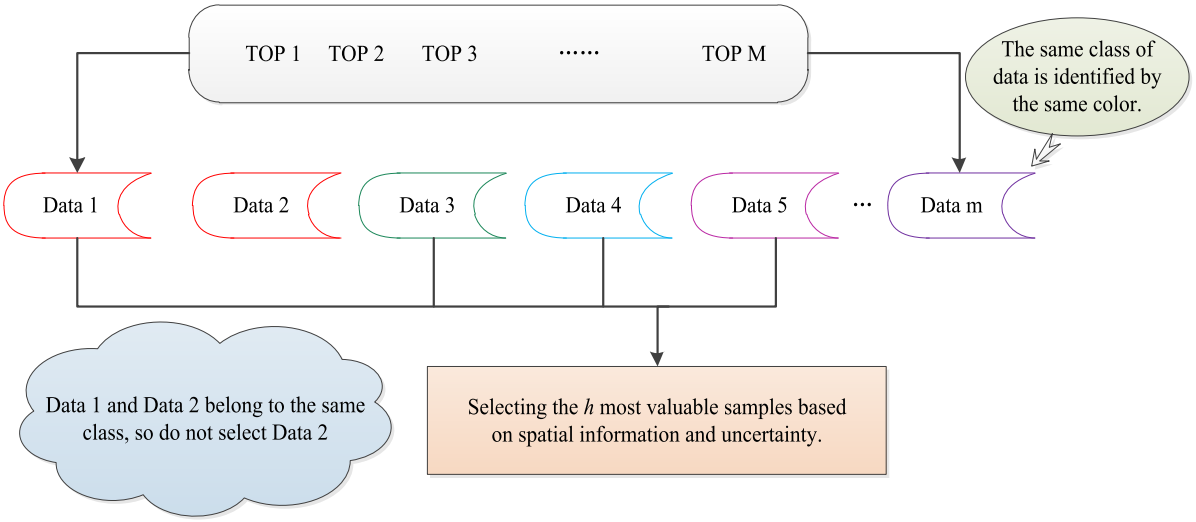

FIGURE 3. Flowchart of the joint selection strategy.

Algorithm 1 Secondary Screening Algorithms

Input: Labeled sample L, unlabeled sample U, T is the limited number of iteration.
While $t < T$:
1.Use L to train Model.
2.Use U as a test set and utilize the Model to classification prediction.
3.Use the active learning strategy to pick out the m most informative samples.
4.Use unsupervised clustering algorithm to cluster m samples, and the number of cluster is h.
5.According to the sample selection strategy, pick out h samples.
6.Label the h samples, updating L and U.
7.Update iteration times $t = t + 1$.
Output: Trained Model

incorrect pseudolabels. Meanwhile, the main advantage of proposed framework is that utilizing a limited labeled samples to effectively train verification classifiers. We can obtain the excellent classification results with the least computational cost.

A. PROPOSED OFSS-SL ALGORITHM

The traditional framework combines single active learning algorithm and semisupervised learning to invoke a collaborative labeling process by both human experts and classifiers. However, applying single active learning algorithm to select unlabeled samples will ignore diversity and discriminative information. Discriminative information and the diversity of samples are vital to improve the accuracy and robustness of the final classifiers. Discriminative information, which represents the quality of the training data, is very important to improve the generalization ability for the classifier. And the diversity and uncertainty of the samples are also as criterion to assess evaluate whether an algorithm can select valuable and representative samples or not. Hence, considering the tradeoff between uncertainty and diversity, we first proposed the one-fold secondary screening algorithm based on semisupervised learning framework (OFSS-SL).

1) DIVERSITY AND DISCRIMINATIVE INFORMATION EXCAVATION

In our experiment, the active learning algorithm is promoted by the unsupervised clustering algorithm. To the best of our knowledge, there are a few works exploited different types of information in the unlabeled data. The traditional methods tend to explore one type of information of the unlabeled data without considering the advantages of various types of information, resulting in information bias. However, secondary screening algorithms can exploit the representative samples that has higher information from different categories.

2) DETAILS OF THE PROPOSED METHOD (OFSS-SL)

Suppose we have a hyperspectral image with n samples $D = \{x_1, x_2, \ldots\ldots, x_n\}$ of d dimensions. We first select m samples from hyperspectral image by applying MCLU, and then cluster them into h groups. We select the most informative samples from each group, that also means we select one representative samples from each category.

After that, we label h samples according to the ground truth as the initial training data $L = \{(x_1, y_1), (x_2, y_2), \ldots\ldots, (x_h, y_h)\}$ with $y_i \in \{1, 2, \ldots\ldots, C\}$ where C is the number of classes in the image. Simultaneously, the labeled data set and the unlabeled data set are updated by adding these newly labeled samples to the labeled set and removing them from the unlabeled set. And the newly labeled samples will train the classifiers. The flowchart and the pseudocode of the OFSS-SL algorithm are illustrated in Fig. 4 and

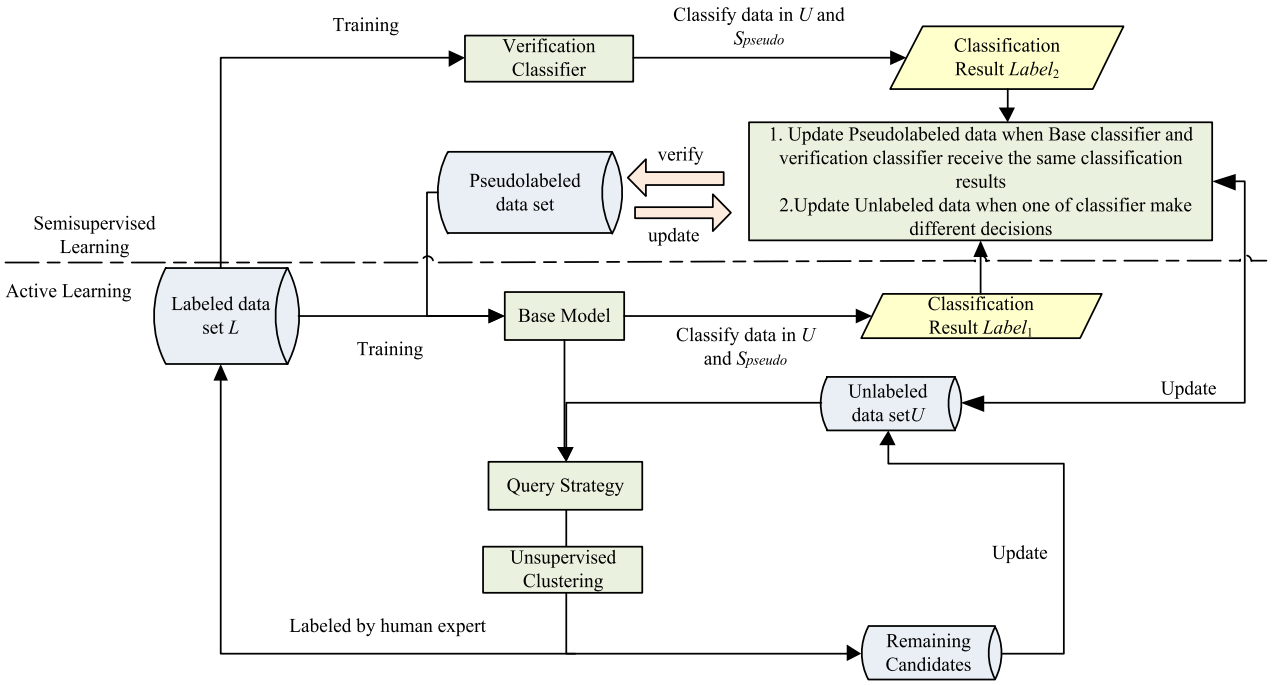

FIGURE 4. Flowchart of the one-fold secondary screening algorithm based on semisupervised learning framework (OFSS-SL).

and Algorithm 2, respectively. Given an initial labeled data set L and an initial unlabeled data set U, the OFSS-SL algorithm starts from an empty set S_{pseudo}, which represents the set of pseudolabeled data. At each iteration, an base classifier is trained using both L and S_{pseudo}. Then, h unlabeled data are selected according to the unsupervised secondary screening algorithms and are labeled by human experts. It should be mention that the h newly labeled samples data will be removed from U and added into L. After that, we utilize a base classifier to classify the samples in both S_{pseudo} and U and obtain result named $Label_1$. Next, we use the expanded labeled data set L to train verification classifier and then applied it to classify the data in both S_{pseudo} and U. Finally, we will obtain two classification results from base classifier and verification classifier, respectively. If the unlabeled data receives the same classification results from two classifiers, it will be assigned with pseudolabels and added into S_{pseudo}. Simultaneously, when the unlabeled samples receive different classification results will be put back into U. The algorithm will terminate until the limit iteration times or becomes an empty set.

B. PROPOSED MSS-MVSL ALGORITHM

The traditional collaborative active and semisupervised learning (CASSL) applies single confidence threshold and AL algorithm for pseudolabeling procedure judgment, which will not be able to meet the changing model. When the initial labeled samples is too few, CASSL may invoke so many wrong pseudolabels and deteriorate the performance of the final classifier. And the effectiveness of classical AL methods relies on whether the unlabeled data that most influence the classification performance can be adopted for human labeling. In order to address these problem, we propose an intensive collaborative framework, a novel multiple-verification semisupervised learning framework

Algorithm 2 OFSS-SL Framework

Input:
Initial training set: $L = \left\{(x_i, y_i)_{i=1}^{l}\right\}$
Initial unlabeled data set: $U = \left(x_j\right)_{j=1}^{u}$
Initial pseudolabeled data set: $S_{pseudo} = \phi$
Initial iteration times: $t = 0$
Query size: h
the number of iterations: T
While $t \leq T$:

1. Repeat:
2. Train base classifier using $S_{pseudo} \cup L$.
3. Applying unsupervised secondary screening algorithm to select h unlabeled samples $U_{Q=}(x_k)_{k=1}^{h}$.
4. Label U_Q by human experts, and update the following two data sets: $L = L \cup U_Q$ and $U = U \backslash U_Q$.
5. Classify $U \cup S_{pseudo}$ using Base Classifier with the classification results being $Label_1$. Classify $U \cup S_{pseudo}$ using Verification Classifier with the classification results being $Label_2$.
6. Update:
   $S_{pseudo} = \{(x_i, Label_1(x_i)) | Label_1(x_i) = Label_2(x_i), x_i \in U \cup S_{pseudo}\}$
   $U = \left\{x_i | Label_1(x_i) \neq Label_2(x_i), x_i \in U \cup S_{pseudo}\right\}$
   iteration times $t = t + 1$.
7. Until T rounds are reached.

Output: Trained Model

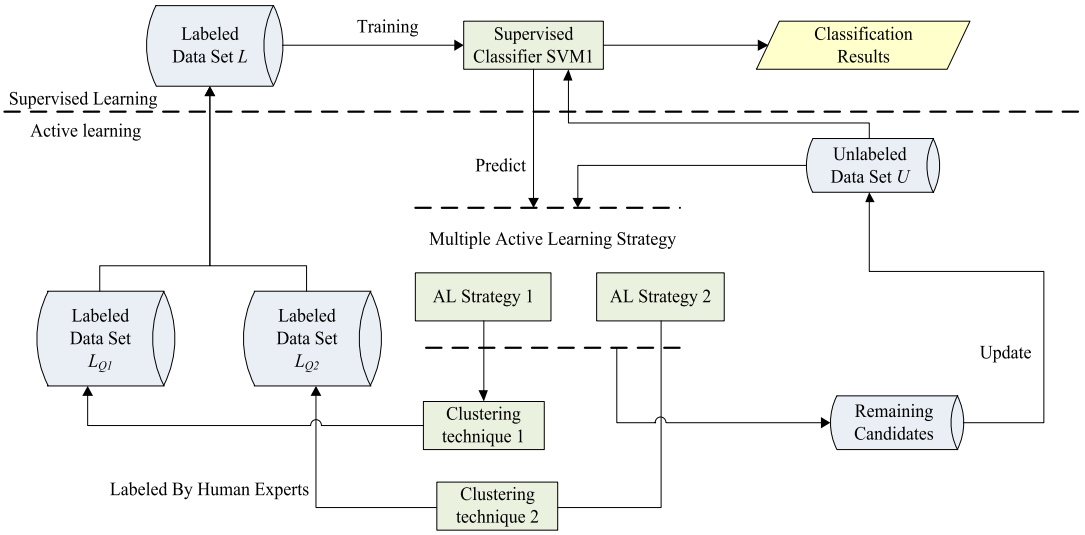

FIGURE 5. Combination of different unsupervised secondary screening algorithm (CUSS).

which combines multiple secondary screening algorithms for guaranteeing the representative and discriminative information.

1) COMBINATION OF SECONDARY SCREENING ALGORITHMS FOR HYPERSPECTRAL IMAGES CLASSIFICATION

We can choose two effective secondary screening algorithms from previous section to obtain valuable samples for manual labeling, respectively. Can we integrate two effective unsupervised secondary screening algorithms? The answer is yes. As an indispensable part in MSS-MVSL, we present an effective framework named combination of unsupervised secondary screening algorithm (CUSS) for joint selecting samples. The idea comes from weighted combination of ensemble learning classifier [65], [66]. The combination of the model is extended to the combination of the strategy so as to achieve the fusion of multiple strategies in a single framework and achieves higher stability. We apply the CUSS to improve the performance of base classifiers, while at the same time as a complementary process to take over the convergence of primary algorithm [59]. The process of the CUSS algorithm is illustrated in Fig. 5 and Algorithm 3. The CUSS significantly improves the generalization abilities of classifiers by applying different unsupervised secondary screening algorithms. Combining all the classification results of multiple classifiers determines the final classification results.

2) DISCRIMINATIVE INFORMATION EXCAVATION AND MULTIPLE VERIFICATION

In the above-mentioned methods, secondary screening algorithms can exploit the representative samples that has higher information from different categories. MSS-MVSL algorithm is proposed with two active learning techniques, two clustering techniques and three classifiers are adopted to assign pseudolabels for the unlabeled data, which can enhance the performance of active learning. Most of the previous studies in active learning process, the generalization ability of classifiers is enhanced gradually at each iteration by adding a set of informative samples to the labeled data set with query function, e.g., MCLU technique. As a result, their effectiveness is highly reliant on adding a set of informative samples and they have a limited ability to excavate discriminative information.

Discriminative information, which represents the quality of the training data, is also vital to improve the generalization ability of the classifier. Because the labeled data is limited. Selecting the unlabeled data as the training data set is an inevitable choice. Therefore, In MSS-MVSL framework, we improve the discriminative information by adding the newly labeled data in the unsupervised secondary screening process and the unlabeled data with pseudolabels. By deploying the unlabeled data with pseudolabels and the labeled data as training data, the discriminative information is enhanced for the classifiers at each iteration to design query function. Another foremost task is that: How to assign pseudolabels to unlabeled data and then retrain the classifiers using both the initially labeled and pseudolabeled data? Deploying an appropriate confidence threshold of the unlabeled sample is a vital step for entire algorithm. On the one hand, a good threshold can guarantee high classification accuracy. While, the high threshold will result in fewer samples satisfying criteria at each iteration, and "empty loop" phenomenon at the beginning of the iteration. On the other hand, a low threshold may result in unlabeled data with low confidence are assigned labels by the classifiers. And the wrong pseudolabels may be implemented into the training procedure and degrade the performance of the final classifier. In addition, the effectiveness of the pseudolabeling procedure also depends on the initial labeled data set. If the initial training set does not match the underlying class distributions, it is difficult to train an

effective classifier at the initial stage and the judgment of unlabeled samples is constantly changing.

In MSS-MVSL, we apply multiple verification model to train two verification classifiers and the discriminative information is improved for the classifiers at each iteration by both the unlabeled data with pseudolabels and the labeled data. Based on the classification results of the three classifiers, we assign the pseudolabels to the unlabeled data if they receive the same labels. Otherwise, we put this samples back to the unlabeled data set.

3) DETAILED STEPS OF MSS-MVSL FRAMEWORK

In this part, we introduce the details of the proposed MSS-MVSL framework. This framework is divided into two parts. The first one is combination of different unsupervised secondary screening algorithm (CUSS). CUSS is the base of the verification part and also reinforces the performance of base classifier. Compared with the single strategies, CUSS deploys the integration of two secondary screening algorithms as joint selection. The other part is double verification to check and classify each unlabeled sample. From the comparison among nine secondary screening algorithms, we select MCLU-KM and nEQB-KM as part in proposed frameworks. We assume that the number of samples to be labeled is Q at each iteration, and the samples contributed by the MCLU-KM is q_1 and the samples contributed by the nEQB-KM strategy is q_2. Designing an appropriate function is a primary task. In this paper, the fitness function of the CUSS is defined as follows.

$$q_1 = Q \times \mathrm{W_A} \tag{9}$$

$$q_1 = Q \times \mathrm{W_B} \tag{10}$$

$$Q = q_1 \mathrm{U} q_2 + R \tag{11}$$

where $\mathrm{W_A}$ and $\mathrm{W_B}$ are weight parameters for MCLU-KM and nEQB-KM, The value of q_1 determines the number of the representative samples selected by MCLU-KM and the value of q_2 determines the number of the representative samples selected by nEQB-KM. We will analyze these parameters in the next section. R is a random factor. $R = Q - q_1 \cup q_2$. If $q_1 \cap q_2 \neq \Phi$, it indicates that the two strategies concurrently select the same most valuable sample. We randomly selected valuable samples with meeting the criteria are treated as supplements. It should be underlined that R does not exist at each iteration, but according to the results of the iteration. Compared with the traditional collaborative active and semisupervised method, MVSL-MSS achieves noticeable improvement. Initially, we utilize the initial labeled samples and the pseudolabels samples to train base classifiers. And base classifier is used to predict the unlabeled data. It should be mentioned that, at the beginning of the iteration, the pseudolabels data is empty. Next, we select q_1 the most informative unlabeled samples by utilizing MLCU-KM technique and select q_2 the most informative unlabeled samples by utilizing nEQB-KM algorithm. And then these unlabeled data are labeled by human experts. We denote q_1 newly labeled samples as L_{Q1} and the q_2 newly labeled samples as L_{Q2}. When the MLCU-KM algorithm and nEQB-KM algorithm select the same sample, we will utilize the random factor R as supplement. Then, these newly labeled data will be removed from the unlabeled set U and added into the labeled set L for training two verifiable classifiers, verification classifier 1 and verification classifier 2. At this time, the two verification classifiers have a large difference in sample distribution due to different strategy. Meanwhile, the unlabeled data with pseudolabels and the unlabeled data are predicted by base model. And we denote the label to the unlabeled data as $Label_1$.

We use L_{Q1} to train verification classifier 1 and use L_{Q2} to train verification classifier 2. Lastly, the unlabeled data with pseudolabels and the unlabeled data are predicted by verification classifier 1 and verification classifier 2, respectively. When base classifier and two verification classifiers obtain the same results on one unlabeled sample, which denotes this result is reliable and then this unlabeled sample will be assigned pseudolabel. On the contrary, when three classifiers have different classification results on a unlabeled sample, this unlabeled sample will be put back to the unlabeled data set for the next iteration. As the algorithm continues to iterate, the performance of the classifiers are improved constantly. Differing form traditional methods, when base classifier and verification classifiers always make same predication on each unlabeled samples, MSS-MVSL exits the part of semisupervised and enter CUSS for next iteration. MSS-MVSL only terminates when it reaches the pre-set iteration times. We believes that three classifiers making the same decision is not the end of the algorithm, it just shows that the algorithm can't obtain more information from unlabeled data with pseudolabels. We still dig out the potential discriminative information and representative information from the unlabeled data. CUSS will be utilized as a backup process to enhance the imperfect end condition. The flowchart of the MSS-MVSL algorithm is illustrated in Figure 6, and the pseudo-code of the MSS-MVSL algorithm is illustrated in Algorithm 3. Lastly, our final combination leads to performance superior to the traditional method, namely CASSL, for remote sensing scene classification.

V. DATASET DESCRIPTION AND DESIGN OF EXPERIMENT

A. DATASET DESCRIPTION

The first hyperspectral data set is Indian pines data set. In June 1992, the NASA AVIRIS image was acquired over the Indian pines agricultural site in northwestern Indiana. AVIRIS is a sophisticated optical sensor system including a number of major subsystems, components, and characteristics [67]. Taking their results in the AVIRIS data characteristics. The AVIRIS sensor receives white light in the foreoptics, disperses the light into the spectrum, converts the photons to electrons, amplifies the signal, digitizes the signal and records the data to high density tape. This data set contains 145 × 145 pixels, at 20-m spatial resolution and 10-nm spectral resolution over the range of 400–2500 nm. Resulting in a

Algorithm 3 MSS-MVSL Framework

Input:

Initial training set: $L = \left\{(x_i, y_i)_{i=1}^{l}\right\}$

Initial unlabeled data set: $U = \left(x_j\right)_{j=1}^{u}$

Initial pseudolabeled data set: $S_{pseudo} = \phi$

Initial iteration times: $K = 0$

Query size: Q

Limit iteration times: T

While $sizeof(U) \geq 10$ and $K \leq T$:

Repeat:

1 Train Base Classifier using $S_{pseudo} \cup L$.

2 Applying Base Model to predict the unlabeled data. Using unsupervised secondary screening strategy A to select q_1 data from $U_{Q1} = (x_{s1})_{s1=1}^{q1}$ and using unsupervised secondary screening strategy B to select q_2 data from $U_{Q2} = (x_{s2})_{s2=1}^{q2}$. Make sure that $Q = q_1 + q_2$.

3 Label $U_Q = U_{Q1} \cup U_{Q2}$ by human experts and update the following four data sets: $L_{Q1} = L \cup U_{Q1}$, $L_{Q2} = L \cup U_{Q2}$, $L = L \cup \left(U_{Q1} \cup U_{Q2}\right)$, $U = U \setminus \left(U_{Q1} \cup U_{Q2}\right)$.

4 Using Base Classifier to Classify $U \cup S_{pseudo}$ and the classification results being $Label_1$

5 Using L_{Q1} to train Verification Classifier 1 and using L_{Q2} to train Verification Classifier 2.

6 Classify $U \cup S_{pseudo}$ by using Verification Classifier 1 and the classification results being $Label_{2\_c1}$, Classify $U \cup S_{pseudo}$ by using Verification Classifier 2 and the classification results being $Label_{2\_c2}$.

7 Update:
$S_{pseudo} = \{(x_i, Label_1(x_i))|Label_1(x_i) = Label_{2\_c1}(x_i) = Label_{2\_c2}(x_i), x_i \in U \cup S_{pseudo}\} U = \left\{x_i|(Label_1(x_i) \neq Label_{2\_c1}(x_i))\text{or}(Label_1(x_i) \neq Label_{2\_c2}(x_i))\text{or}(Label_{2\_c1}(x_i) \neq Label_{2\_c1}(x_i))\right\}$, $x_i \in U \cup S_{pseudo}$

8 $K = K + 1$

9 While $sizeof(U) < 10$ and $K \leq T$:

10 Run:
Algorithm 3 CUSS Framework

Return: Trained Model

TABLE 1. Numbers of samples for the corresponding classes of the Indian Pines data set.

Class Label	Class Name	#samples
C1	Corn-no till	1428
C2	Corn-min till	830
C3	Corn	237
C4	Grass/Pasture	483
C5	Grass/Trees	730
C6	Hay-windrowed	478
C7	Soybeans-no till	972
C8	Soybeans-min till	2455
C9	Soybean-clean till	593
C10	Wheats	205
C11	Woods	1265
C12	Building-Grass-Tress-Drives	386

TABLE 2. Numbers of samples for the corresponding classes of the kennedy space center data set.

Class Label	Class Name	#sample
C1	Scrub	761
C2	Willow	243
C3	CP Hammock	256
C4	CP/Oak	252
C5	Slash Pine	161
C6	Oak/Broadleaf	229
C7	Hardwood Swamp	105
C8	Graminiod Marsh	431
C9	Spartina Marsh	520
C10	Cattail Marsh	404
C11	Salt Marsh	419
C12	Mud Floats	503
C13	Water	927

TABLE 3. Numbers of samples for the corresponding classes of the Pavia University data set.

Class Label	Class Name	#sample
C1	Asphalt	6631
C2	Meadows	18649
C3	Gravel	2099
C4	Trees	3064
C5	Painted metal sheets	1345
C6	Bare Soil	5029
C7	Bitumen	1330
C8	Self-Blocking Bricks	3682
C9	Shadows	947

200-band image, twenty noisy and water absorption bands (104–108, 150–163, and 220) were removed [59]. It should be underlined that we select 12 categories that the number of land cover samples more than 100 for classification. And 10062 available samples.

The second hyperspectral data set is Kennedy Space Center (KSC), which was acquired over the Kennedy Space Center, Florida, on March 23, 1996, at a spatial resolution of 18 m. The original data set consists of 220 bands, and it has a size of 512 × 614 pixels, after removing water absorption and low SNR bands, 176 bands are left. The available data were collected using land-cover maps derived from color infrared photography provided by KSC and Landsat thematic mapper imagery. This data set contains 13 classes. A total of 5211 pixels are labeled with the different classes of land cover [35].

Pavia university data set was acquired by the Reflective Optics System Imaging Spectrometer (ROSIS) instrument in 2001, covering the city of Pavia, Italy [68]. The principle and performance of ROSIS will be presented in the following with emphasis on the sensor, signal conditioning and the related data flow. The panchromatic information of a ground track with several elements is imaged on a

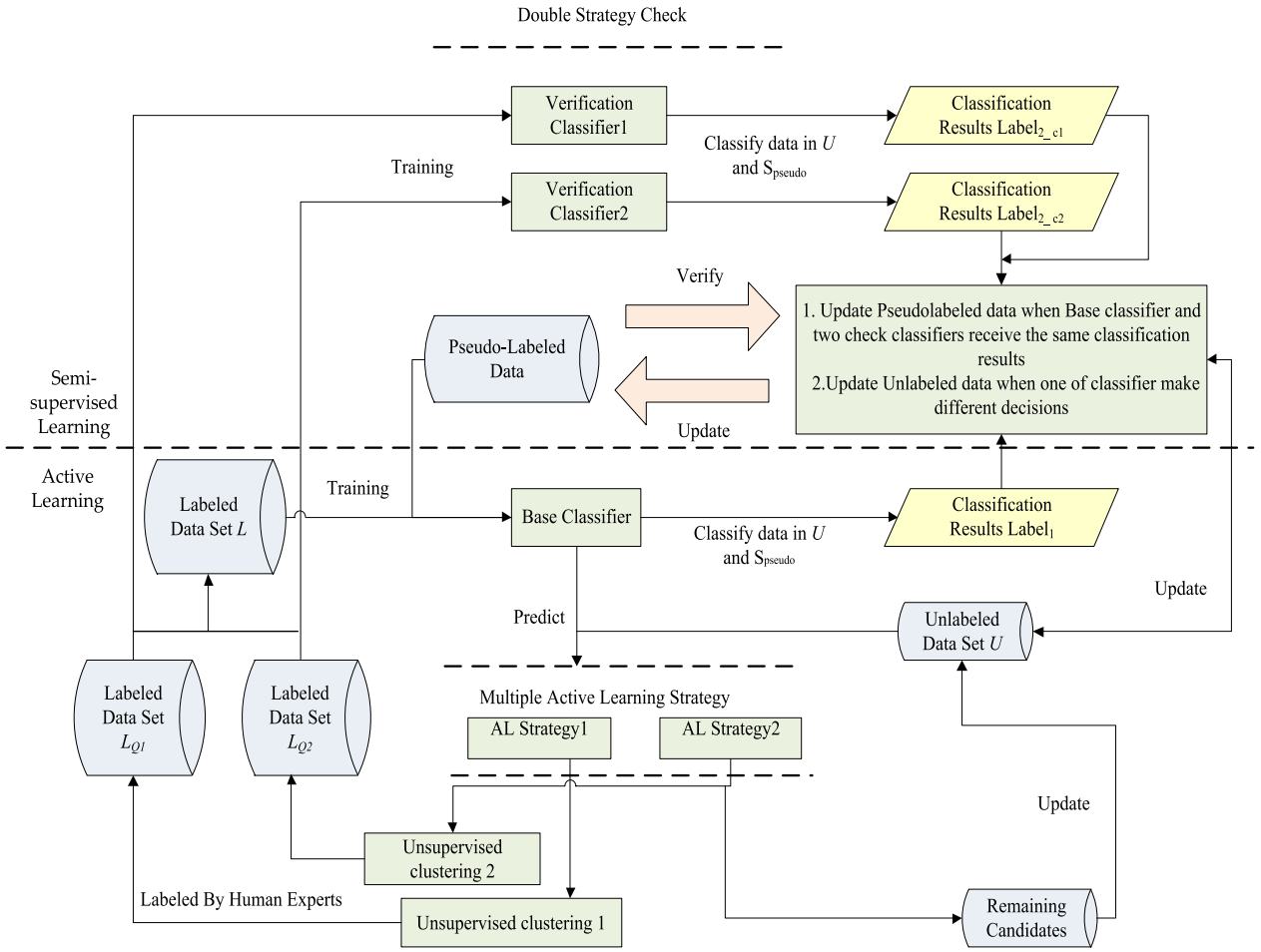

FIGURE 6. Flowchart of the syncretic multiple secondary screening algorithms and multi-verification semisupervised learning framework (MSS-MVSL).

matrix CCD sensor. One line of the track corresponds to one CCD line. The panchromatic light of this line is diffracted by means of a grating. The spectral information is represented by the different CCD lines. The spatial resolution of each single ground element in flight direction is derived from the ground orbit velocity multiplied by the exposure time. The spectral resolution depends primarily on the grating and on the imaging optics. But the sensitivity and characteristics of the imaging spectrometer depend on the selection of the appropriate CCD matrix sensor. The image scene is centered at the University of Pavia, with a size of 610 × 340 pixels. It comprises 115 spectral channels in the wave-length range from 0.43 to 0.68 um with a spatial resolution of 1.3 m, and 103 channels are used in the experiment after noise and water absorption bands are removed. This data set contains nine classes representing the different types of land cover, and there are 42776 available samples [59].

B. DESIGN OF EXPERIMENT

1) EXPERIMENTAL SETUP ON UNSUPERVISED SECONDARY SCREENING ALGORITHMS

For Indian pines data set, we selected 12 categories for classification, and has 10062 labeled pixels (the number of sample more than 100). Firstly, we randomly divided the total available data into two parts for each image: 70% data sets used for training and 30% for testing. And for the 70% training data, we randomly selected five samples in each class as the initial labeled data, and then remaining samples were used as the unlabeled data for the unsupervised secondary screening. For KSC data set, we randomly divided the total available data into two parts: 50% for training and 50% for testing. And the training data contains the unlabeled samples and the labeled samples. We randomly selected five samples from the each large sized class (the number of samples more than 200) as the initial labeled data. We selected three labeled samples from slash pine and selected two labeled samples from hardwood swamp as the initial labeled data. For Pavia university data set, we randomly divided the total available data into two parts: 75% for training and 25% for testing. We randomly selected thirty samples in each class as the initial labeled data.

In order to provide some guidelines to the users under different conditions. We assess the compatibility of nine considered secondary screening algorithms on different data sets. At each iteration, 10 samples were selected by manual labeling, and added to the labeled data set. Each algorithm iterates

TABLE 4. Performance comparison showing OA obtained from nine algorithms on the indian pines data set.

Number of labeled samples	MCLU-KM (%)	MCLU-SC (%)	MCLU-HC (%)	nEQB-KM (%)	nEQB-SC (%)	nEQB-HC (%)	aEQB-KM (%)	aEQB-SC (%)	aEQB-HC (%)
200	72.16	71.86	71.75	71.48	72.04	71.94	70.62	72.64	71.53
350	77.76	77.61	76.67	77.05	77.04	77.14	75.93	76.58	76.32
500	80.65	80.50	80.13	79.63	79.46	80.26	79.07	79.02	79.42
650	82.12	82.16	81.50	81.67	81.83	82.08	81.58	81.08	80.92
800	83.30	82.35	83.46	83.15	83.10	83.60	83.59	83.09	82.92
950	84.14	83.54	84.46	84.58	84.02	84.37	84.53	84.12	84.13

TABLE 5. Performance comparison showing AA obtained from nine algorithms on the indian pines data set.

Number of labeled samples	MCLU-KM (%)	MCLU-SC (%)	MCLU-HC (%)	nEQB-KM (%)	nEQB-SC (%)	nEQB-HC (%)	aEQB-KM (%)	aEQB-SC (%)	aEQB-HC (%)
200	71.27	69.45	70.69	68.01	68.66	68.25	68.29	69.30	69.16
350	76.99	75.63	76.01	74.27	74.14	74.78	74.38	73.56	74.71
500	80.10	79.32	80.03	77.27	76.75	78.36	78.11	77.00	78.68
650	81.89	81.35	81.40	80.03	79.69	80.74	81.24	79.49	80.48
800	82.92	82.35	83.15	82.12	81.60	82.49	83.31	82.05	82.64
950	83.94	83.54	84.38	84.03	82.94	83.57	84.34	83.54	83.90

TABLE 6. Prformance comparison showing OA and STD obtained from nine algorithms on the KSC data set.

Number of labeled samples	MCLU-KM (%)	MCLU-SC (%)	MCLU-HC (%)	nEQB-KM (%)	nEQB-SC (%)	nEQB-HC (%)	aEQB-KM (%)	aEQB-SC (%)	aEQB-HC (%)
200	89.79	89.91	90.56	89.63	88.93	89.53	89.73	89.21	89.68
350	92.94	92.57	92.91	92.08	91.60	92.56	92.33	91.97	92.40
500	93.72	93.41	93.76	93.43	92.82	93.32	93.33	93.14	93.47
650	93.92	93.93	93.96	93.76	93.38	93.75	93.75	93.42	93.71
800	94.12	94.11	94.11	93.89	93.69	93.92	93.84	93.82	93.77
950	94.14	94.14	94.15	93.96	9372	93.91	93.95	93.78	93.86

95 times, We have conducted this experiment about 20 times for each algorithm. For all the compared methods, the adopted classifier was support vector machine (SVM) classifiers based on the radial-basis-function kernel, SVM is a supervised nonparametric statistical learning technique, which is not constrained to prior assumptions on the distribution of the input data. There are two parameters for the SVM classifier, i.e., the Gaussian kernel parameter G and the penalty coefficient C, which are usually selected via cross validation. $C = \{2^{-5}, 2^{-3}, \ldots\ldots, 2^{15}\}$ is the hunting zone for penalty coefficient, $G = \{2^{-15}, 2^{-13}, \ldots\ldots, 2^{3}\}$ is the hunting zone for Gaussian kernel parameter. The sample selection method is TopN. The experimental simulation environment is Inter(R) Core(TM) i7-6700HQCPU@2.6Ghznotebook, and its memory is 16G, and the operating system is Win10. Using the python scikit-learning algorithm package to simulate the experiment.

2) EXPERIMENTAL SETUP ON PROPOSED METHODOLOGY

With regard to validate the effectiveness of two proposed frameworks, we presents experiments on original collaborative active and semisupervised learning (CASSL), nEQB-KM, MCLU-ECBD algorithm with a comparative study. In the experiments, for every algorithm, twenty runs were executed on each data set with different initial labeled data. For all the compared methods, the adopted classifier was SVM based on the radial-basis-function kernel. OFSS-SL, MSS-MSAL are batch-mode active-learning algorithm. This will bring two additional parameters, the candidate query set with m samples is selected by the uncertainty, and the actual query set with h samples is selected from the the candidate query set by the diversity. In our experiments, we set $h = 10$ and $m = 40$.

C. RESULTS AND ANALYSIS

1) RESULTS ON COMPARISON AMONG SECONDARY SCREENING ALGORITHMS

In order to analyze the effectiveness of the investigated techniques, we decided to perform re-estimation for nine considered secondary screening algorithms. (MCLU-KM, MCLU-SC, MCLU-HC, nEQB-KM, nEQB-SC, nEQB-HC, aEQB- KM, aEQB-HC, aEQB-SC). Tables 4–5 shows the classification performance on the Indian pines data set. Tables 8 compares the running time of the all the investigated secondary screening algorithms. In the Indian pines data set. MCLU-KM has the better performance than others.

TABLE 7. Performance comparison showing AA and STD obtained from nine algorithms on the KSC data set.

Number of labeled samples	MCLU-KM (%)	MCLU-SC (%)	MCLU-HC (%)	nEQB-KM (%)	nEQB-SC (%)	nEQB-HC (%)	aEQB-KM (%)	aEQB-SC (%)	aEQB-HC (%)
200	85.00	85.68	86.15	84.55	84.33	83.75	84.95	84.25	85.11
350	89.27	89.01	89.36	87.79	87.77	88.47	88.25	88.22	88.53
500	90.39	90.04	90.42	89.97	89.34	89.92	89.86	89.73	90.15
650	90.79	90.68	90.80	90.52	90.06	90.42	90.45	90.08	90.39
800	91.05	91.00	91.04	90.62	90.43	90.62	90.51	90.51	90.45
950	91.06	91.05	91.06	90.76	90.45	90.68	90.74	90.48	90.62

TABLE 8. Total training time (in seconds) of nine compared algorithms on the KSC data set and indian pines data set.

Algorithm	Total time(s) (Indian pines data set)	Total time (s) (Ksc data set)
MCLU-KM	380.45	266.24
MCLU-SC	382.02	261.13
MCLU-HC	379.17	270.92
nEQB-KM	1467.13	805.46
nEQB-SC	1517.24	814.57
nEQB-HC	1463.45	807.45
aEQB-KM	1590.18	849.45
aEQB-SC	1668.22	814.05
aEQB-HC	1595.11	799.25

In addition, aEQB-KM achieve the highest accuracy when the algorithms finish iterating and MCLU-HC needs the least time for iterating. Tables 6–7 show the classification performance on the KSC data set, as for aEQB, both aEQB-HC and aEQB-KM have better performance on OA, AA and Kappa. As for nEQB, nEQB-KM is the most efficient and has the best classification performance. As for MCLU, MCLU-HC consistently demonstrates statistically better performance when compared to each of the other two algorithms. However, MCLU-HC spends a little more time than other algorithms. From this comparison, we select MCLU-KM and nEQB-KM as a part of proposed frameworks.

2) RESULTS ON PROPOSED METHODOLOGY

Taking into account the aforementioned method, for experimental parameters W_A and W_B, ($W_B = 1 - W_A$), a range of parameters were considered ($W_A = 0.1, W_A = 0.3, W_A = 0.5, W_A = 0.7, W_A = 0.9$). From Fig. 11, we can observe that $W_A = 0.5$ almost performs better than other parameters at each iteration. This phenomenon can be attributed that complementing nEQB-KM and MCLU-KM for each other will stimulate the performance of overall methods. Hence, we simply split the data as $q_1 = q_2 = Q/2$.

And set the value of W_A and W_B are the same, both of them are 0.5. Then, these newly labeled data will be removed from U and add into L. The quantitative evaluations of the three data sets are shown in tables 10–12, with overall accuracy (OA) standard deviation (STD), training time and paired t-test at 95% significance level with p-value. T-test is widely used to verify the difference between two data sets, which can suggest which method is better in the whole

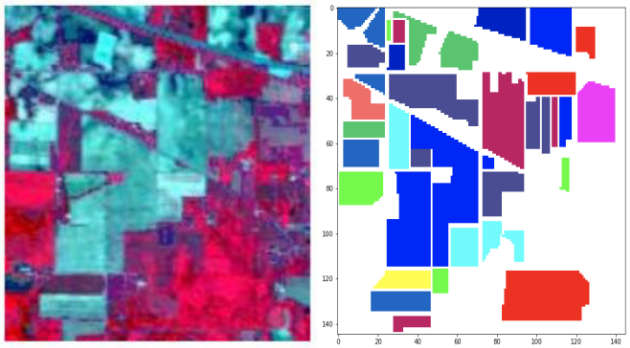

FIGURE 7. False-color composite image of Indian Pines data set and color map of ground truth. (a) False-color image. (b) Ground truth.

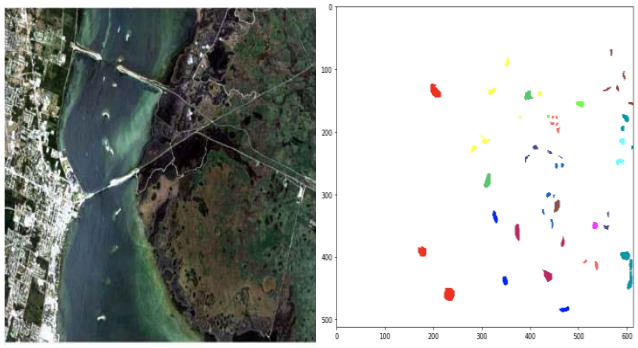

FIGURE 8. False-color composite image of KSC data set and color map of ground truth. (a) False-color image. (b) Ground truth.

experiments. We regarded the results of MSS-MVAL and OFSS-SL as the first data set, repesectively. And the other results of each compared method were regarded as the second data set. Hence, we obtain two p-values. We donate the p-value obtained by OFSS-SL as p-value 1. We donate the p-value obtained by MSS-MVSL as p-value 2. If the p-value obtained by this test is smaller than or equal to 0.05, which can be regarded that this algorithm with the better performance in this comparison. Both p-value 1 and p-value 2 confirm that OFSS-SL and MSS-MVSL perform better than or similar to the competitors.

The principle of active learning is to use fewer labeled samples to get the better training effect. Therefore, the number of labeled samples represents the labor cost and measures the consumption of active learning in the iterative phase. Because

TABLE 9. Proposed methods versus the comparison methods based on paired t-tests at 95% significance level, and a performance comparison showing the OA and STD obtained with the indian pines data set.

Number of labeled samples	nEQB-KM AVG%±STD%	CASSL AVG%±STD%	MCLU-ECBD AVG%±STD%	OFSS-SL AVG%±STD%	MSS-MVSL AVG%±STD%
200	68.58±1.28	66.89±1.50	67.39±1.00	68.85±0.42	69.93±0.31
350	74.65±1.23	74.01±2.01	73.31±1.19	75.89±1.20	76.65±1.22
500	78.79±0.71	78.17±1.53	77.37±0.45	78.90±1.50	80.55±0.80
650	81.05±0.43	81.45±1.15	81.57±0.59	81.86±0.92	82.27±0.79
800	82.43±0.55	83.01±0.60	83.05±0.64	83.34±1.00	84.29±0.60
950	83.93±0.40	84.01±0.72	84.00±0.47	84.58±0.56	84.72±0.61

Number of labeled samples	nEQB-KM		CASSL		MCLU-ECBD	
	p-value 1	p-value 2	p-value 1	p-value 2	p-value 1	p-value 2
200	0.47	0.032	2.81e-05	1.10e-05	2.35e-08	6.81e-07
350	3.57e-04	2.62e-07	1.10e-04	4.09e-07	2.03e-13	1.06e-16
500	0.73	2.90e-10	0.048	2.54e-09	4.84e-05	8.89e-22
650	8.72e-05	6.71e-08	0.11	1.20e-05	3.52e-10	1.5e-04
800	8.41e-04	7.95e-14	0.017	3.84e-08	1.08e-14	1.16e-11
950	3.39e-05	1.87e-06	2.48e-03	1.12e-04	3.45e-07	4.77e-04

TABLE 10. Total training time (in seconds) of five compared algorithms on the three data set.

Algorithm	Iterate 90 times time (s) (Indian pines data set)	Iterate 25 times time (s) (Ksc data set)	Iterate 90 times time (s) (Pavia University data set)
MCLU-ECBD	480.55	106.24	1748.89
CASSL	4816.27	329.31	32791.27
nEQB-KM	1442.07	261.13	4406.57
OFSS-SL	4030.96	270.92	32081.43
MSS-MVSL	4017.13	238.26	28295.90

of the technique of batch extraction, the minimum unit cost of manual labeling is h, and it is the number of batch samples. The cost of manual labeling as follows:

$$\begin{aligned} COST &= (N+1) \times h - \frac{1}{2}h \\ &= Nh + \frac{1}{2}h, P(N) < p < P(N+1) \end{aligned} \tag{12}$$

where $P(N)$ is the model precision at Nth iteration, the value of p can be specified as OA, AA or Kappa coefficient, respectively.

For human-labeling effort, with respect to the original methods, significant decreasing in human labeling are achieved by two proposed methods. From Tables 9–11, we can observe that when all the algorithms reach the certain OA value, the proposed algorithms need less labeling costs. In Indian pines data set, we can find that proposed frameworks consistently demonstrate statistically better performance when compared to each of the other three algorithms. From table 9, we can observe that CASSL has worse classification performance at initial stage, which is known as the "old start" problem. This is because too few initial labeled samples will lead to many errors obtained by inaccurate classifiers at the initial training stage and also impact the final classification performance. In KSC data set, CASSL astringes too early. The accuracy of CASSL on the KSC data set after 250 samples maintains the status quo. With regard to the verification strategy, the difference of model directly decided the reliability of the pseudolabels. Moreover, the stopping condition of the CASSL framework heavily relies on the reliability of the verification procedure, which will result in CASSL framework end so early. From 12 (c), we can observe that OFSS-SL may have similar performance as MSS-MVSL. But the training time of MSS-MVSL is more shorter. Hence, choosing MSS-MVSL is unavoidable selection.

FIGURE 9. False-color composite image of Pavia University data set and color map of ground truth. (a) False-color image. (b) Ground truth.

TABLE 11. proposed methods versus the comparison methods based on paired t-tests at 95% significance level, and a performance comparison showing the OA and STD obtained with the pavia university data set.

Number of labeled samples	nEQB-KM AVG%±STD%	CASSL AVG%±STD%	MCLU-ECBD AVG%±STD%	OFSS-SL AVG%±STD%	MSS-MVSL AVG%±STD%
400	87.22±1.31	87.31±0.57	86.69±0.87	87.45±0.95	87.67±0.37
500	89.12±0.65	89.37±0.40	88.23±0.89	89.88±0.23	89.87±0.12
600	91.02±0.31	90.97±0.31	89.51±0.62	91.20±0.14	91.45±0.08
700	91.92±0.36	91.84±0.42	91.61±1.72	91.96±0.14	92.21±0.11
800	92.55±0.50	92.39±0.24	91.88±2.42	92.72±0.25	92.76±0.20
900	92.86±0.44	92.87±0.26	92.32±2.14	92.89±0.15	93.35±0.26
1000	93.22±0.48	93.30±0.07	92.81±1.05	93.52±0.05	93.64±0.11
1100	93.27±0.01	93.32±0.06	93.01±0.80	93.67±0.13	94.05±0.33

Number of labeled samples	nEQB-KM		CASSL		MCLU-ECBD	
	p-value 1	p-value 2	p-value 1	p-value 2	p-value 1	p-value 2
400	0.37	0.163	0.021	0.046	9.03e-03	8.11e-06
500	3.09e-06	2.37e-07	0.016	0.85	2.89e-04	8.18e-04
600	2.98e-03	1.77e-04	9.72e-04	0.37	7.03e-03	6.89e-04
700	0.45	9.93e-04	0.41	0.71	0.20	0.031
800	0.042	0.035	0.014	0.02	0.032	0.032
900	0.19	4.01e-05	0.71	8.16e-07	0.025	1.98e-04
1000	7.56e-05	1.89e-05	4.28e-03	1.34e-11	0.036	0.035
1100	6.89e-06	4.03e-05	9.23e-05	1.23e-09	3.12e-05	8.96e-04

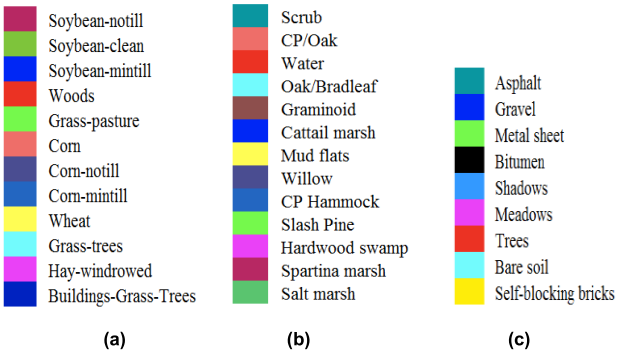

FIGURE 10. Class legends for three data sets. (a)Indian Pines. (b) KSC. (c) Pavia University.

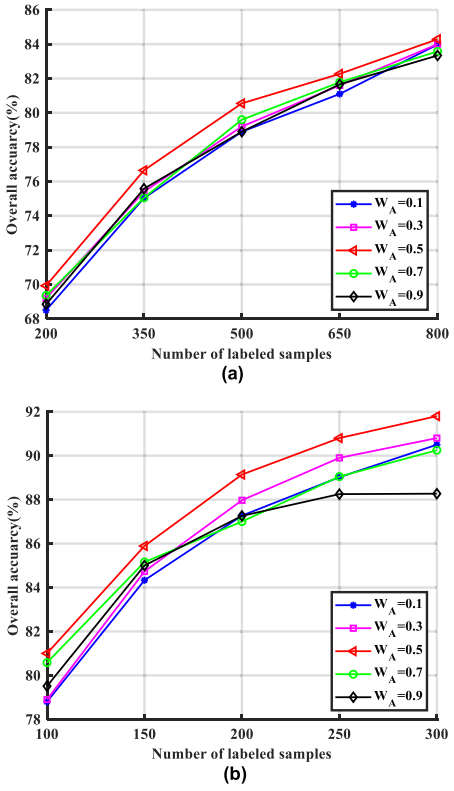

FIGURE 11. The various value of W_A utilized in unsupervised secondary screening algorithm for (a) Indian pines dataset, and (b) KSC data set.

For the time cost, all of the algorithms in our experiments implemented python scikit-learning package. Moreover, compared with CASSL, we can find that MSS-MVSL and OFSS-SL are more effective from Table 10. Generally speaking, MSS-MVSL and OFSS-SL may need extra time to mine the representative information in the unlabeled data. However, as a parallel programming framework, MSS-MVSL can exploit multiple architectures. It generates a set of parallel representative samples selected form different algorithms and assigns label to each one at each iterations. By clustering the informative unlabeled samples, OFSS-SL will find the most representative more quickly and accurately.

In conclusion, the proposed framework can use less time to achieve better performance. Fig. 13 shows that the comparison of the final classification map on the Indian pines data set, we can apparently observe that OFSS-SL and MSS-MVSL have better classification results.

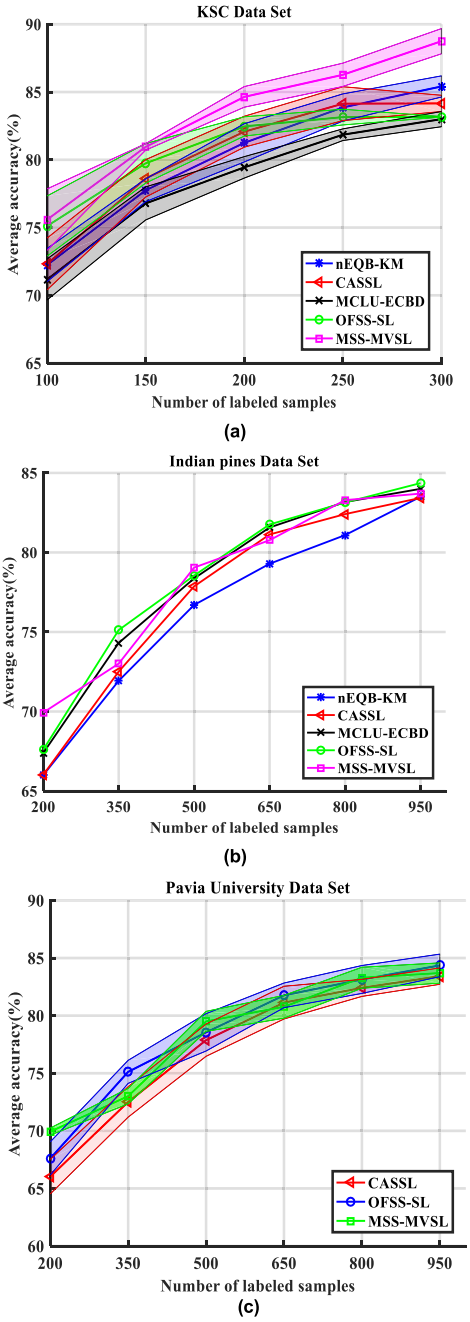

FIGURE 12. Average classification results of the different algorithms on the two hyperspectral images (a) KSC data set. (b) Indian Pines data set. (c) Pavia University data set.

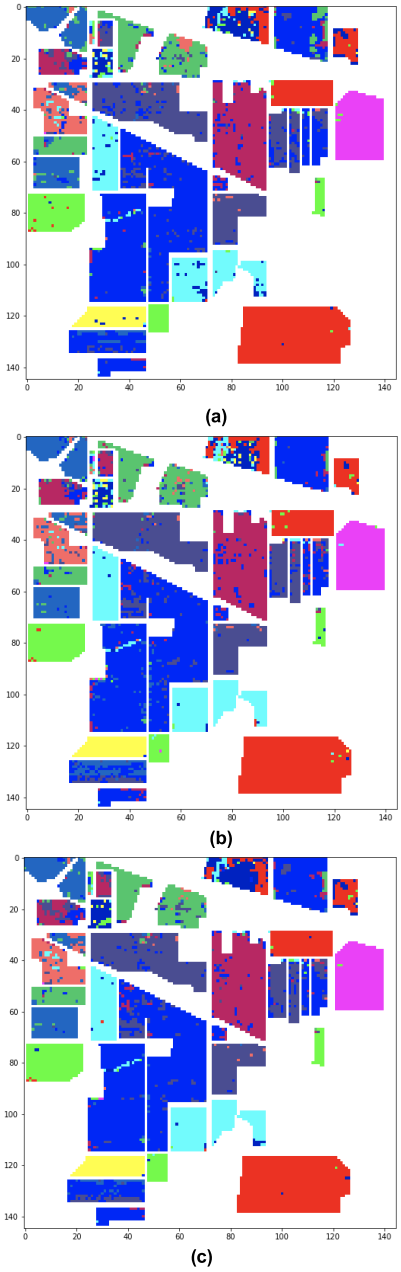

FIGURE 13. Comparison of the final classification map of different framework on the Indian Pines dataset. (a) CASSL; (b) OFSS-SL; (c) MSS-MVSL.

3) PARAMETER SENSITIVITY ANALYSIS

We have performed a sensitivity analysis of the parameters m and h in OFSS-SL, MSS-MVSL. The parameter m controls the number of unlabeled pixels to be selected and h represents the number of unlabeled samples to be labeled by human experts at each iteration. This paper has set different m values (20, 40) and h values (10) to conduct OFSS-SL, MSS-MVSL on the three data sets, respectively.

We design three groups of value ($m = 40, h = 10$; $m = 40, h = 20$; $m = 20, h = 10$). We analyze these parameters by undertaking experiments on the Indian pines data set. Fig. 14 shows the results with different specific parameter values over ten runs. We can observe that OFSS-SL and MSS-MVSL are sensitive to the batch size. For MSS-MVSL, when we set $m = 40, h = 10$, it can obtain a better performance. It can be noted that when the value of h are fixed, $m = 20$, MSS-MVSL will finish iterating more quickly. When the value of m is same, batch size h will determine the running time of the framework. The number of h is lager, the computational time is longer. For, MSS-MVSL $m = 40, h = 10$ is also a more appropriate parameter settings. However, form the figure 12 (a), with the

TABLE 12. proposed methods versus the comparison methods based on paired t-tests at 95% significance level, and a performance comparison showing the OA and STD obtained with the KSC data set.

Number of labeled samples	nEQB-KM AVG%±STD%	CASSL AVG%±STD%	MCLU-ECBD AVG%±STD%	OFSS-SL AVG%±STD%	MSS-MVSL AVG%±STD%
100	78.74±1.27	76.70±2.25	77.69±1.01	79.51±2.58	81.00±2.14
150	84.26±0.77	83.02±1.38	82.46±0.92	85.12±1.17	85.89±0.62
200	87.27±0.60	86.70±1.27	85.89±0.69	87.26±0.98	89.14±0.43
250	89.02±0.77	88.43±1.37	87.89±0.39	88.25±0.35	90.80±0.51
300	90.48±0.31	88.43±0.01	88.00±0.42	88.27±0.01	91.80±0.46

Number of labeled samples	nEQB-KM		CASSL		MCLU-ECBD	
	p-value 1	p-value 2	p-value 1	p-value 2	p-value 1	p-value 2
100	0.021	6.50e-06	1.65e-06	1.40e-08	5.24e-07	4.90e-10
150	0.047	8.48e-11	2.99e-09	1.67e-11	1.53e-05	9.45e-18
200	0.11	1.90e-15	0.067	3.71e-10	8.87e-05	4.45e-21
250	2.58e-06	1.89e-11	0.010	2.16e-08	4.45e-06	6.42e-26
300	2.55e-27	2.09e-13	1.34e-05	3.64e-13	1.18e-05	2.68e-30

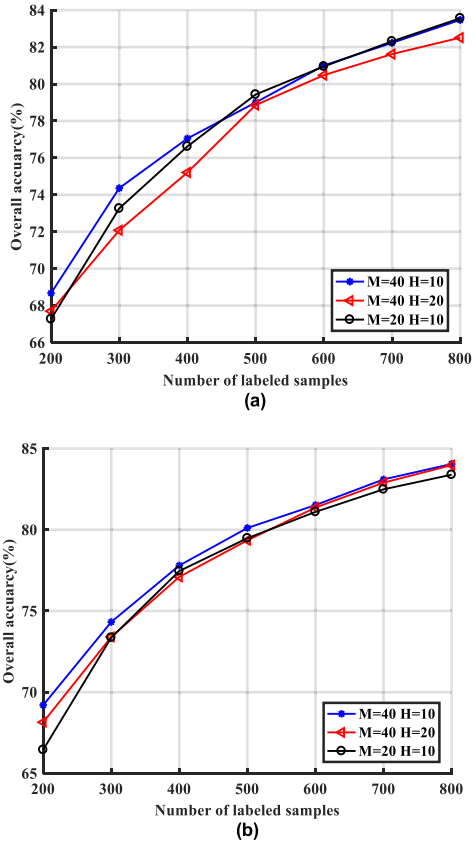

FIGURE 14. Overall classification accuracy versus the number of training samples obtained by the OFSS-SL and MSS-MVSL framework with different m and h values for Indian pines data sets. (a) OFSS-SL; (b) MSS-MVSL.

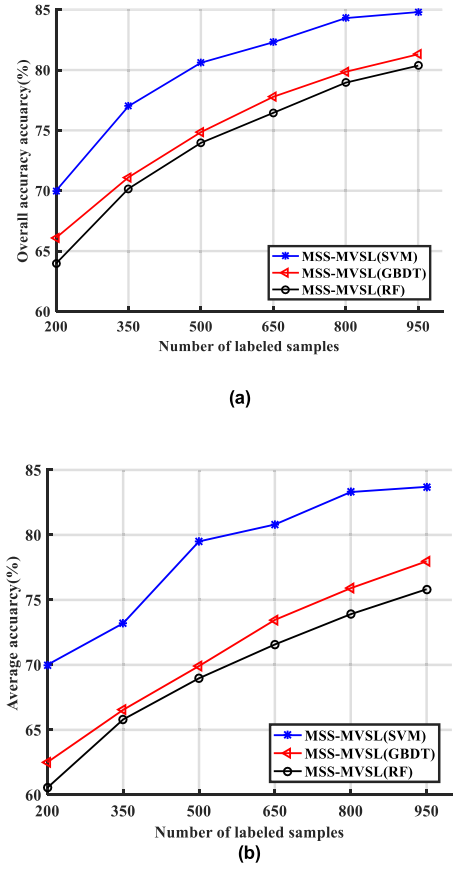

FIGURE 15. Classification accuracy versus the number of training samples obtained by the MSS-MVSL technique with different verification classifiers for Indian pines data sets. (a) Overall accuracy. (b) Average accuracy.

increasing of iterations, when set $m = 20$, $h = 10$, OFSS-SL also can obtain a promising results.

4) FURTHER ANALYSIS

All of the algorithms in our experiments implemented the same classifiers (SVM). In MSS-MVSL, we apply two verification classifiers and one base classifier and all of them are SVM. However, it is important that the classifiers considered in the module should be diverse, and their performance should be complementary. Hence, we investigate whether classifiers potentially contribute to improve the final classification performance.

TABLE 13. Computational Time of Labeling 900 unlabeled samples (in seconds) Taken from OFSS-SL and MSS-MVSL Framework with Respect to Different m and h Values on Indian pines data set.

Parameter settings	Total time(s) OFSS-SL	Total time(s) MSS-MVSL
M=40, h=10	4030.96	4017.13
M=40, h=20	2592.82	2238.54
M=20, h=10	3979.13	3968.42

TABLE 14. Computational time of labeling 900 unlabeled samples (in seconds) taken from three compared algorithms on the indian pines data set.

Algorithm	Total time(s) (Indian pines data set)
MSS-MVSL(GBDT)	6173.36
MSS-MVSL(RF)	3699.15
MSS-MVSL(SVM)	4017.13

We apply Random forest (RF) [70] and Gradient Boosting Decision Tree (GBDT) [71] as verification classifiers, respectively. RF is a classifier constructed from an ensemble of classification and regression trees (CART). It consists of a combination of classifiers where each classifier contributes with a single vote for the assignation of the most frequent class to the input vector (x). Moreover, RF also can handle a high-dimensional feature space with less computation, but it is insensitive to noise in training samples. By contrast, GBDT is a gradient boosting algorithm that utilizes decision stumps or regression tress as weak classifiers, and it is generated serially, which is sensitive to abnormal data. In terms of the previous classification results, we set $m = 40, h = 10$ in this group test and the model parameters of RF and GBDT are the same. The results obtained on the Indian data sets are shown in Fig. 14. We donate MSS-MVSL with two SVM verification classifiers as MSS-MVSL(SVM), donate MSS-MVSL with two GBDT verification classifiers as MSS-MVSL(GBDT) and donate MSS-MVSL with two RF verification classifiers as MSS-MVSL(RF). For every framework, ten runs were executed.

We can observe that MSS-MVSL(SVM) consistently outperforms MSS-MVSL(GBDT) and MSS-MVSL(RF). From the table 14, we also observe that the largest computational time is obtained with MSS-MVSL(GBDT). Although the smallest computational time is obtained with MSS-MVSL(RF), the accuracy of MSS-MVSL(RF) is the lowest. Hence, we apply SVM classifiers as verification classifiers is an promising choose.

VI. CONCLUSION AND FUTURE WORK

We proposed two novel algorithms OFSS-SL, MSS-MVSL for hyperspectral image classification, which combine secondary screening algorithms and semisupervised learning in a collaborative manner. In order to choose better secondary screening algorithms for further study, we first investigated different secondary screening algorithms on two data sets. Then, we chose two different secondary screening algorithms with excellent performance and implemented them in our proposed framework. OFSS-SL exploits both the diversity and representative information by applying unsupervised secondary screening, meanwhile, excavating discriminative information by assigning pseudolabels to the unlabeled data, thereby gradually improving the classification performance with the labeled data and the unlabeled data with pseudolabels. The main novelty of MSS-MVSL lies in its multiple verification and the integration of multiple unsupervised secondary screening algorithms, which aims to ensure the quality of pseudolabels and complements each other. Differing from the the traditional combination methods, we consider the complementary characteristics of the secondary screening algorithms and mine the multiple validation information, which makes the proposed technique more robust at the initial stage. We have compared them with three state-of-the-art methods by using three real hyperspectral data sets. By comparison, we can observe that for all the considered data sets, the proposed techniques consistently provide better accuracy and stability.

In this framework, we utilize multiple check models will cause multiple consumptions, and we apply a parallel computing framework to simultaneously train model. Hence, an important issue deserves to further investigate. How to reduce computational cost when improve the verification procedure? We will utilize such as Spark MLlib framework. Reference [72] and Apache Mahout technique. Reference [73] to migrate the computational costs in the future. We also plan to incorporate novel AL techniques into our framework.

REFERENCES

[1] S. Yang, Z. Shi, and W. Tang, "Robust hyperspectral image target detection using an inequality constraint," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 6, pp. 3389–3404, Jun. 2015.

[2] S. Yang and Z. Shi, "Hyperspectral image target detection improvement based on total variation," *IEEE Trans. Image Process.*, vol. 25, no. 5, pp. 2249–2258, May 2016.

[3] H. Pu, Z. Chen, B. Wang, and G. Jiang, "A novel spatial-spectral similarity measure for dimensionality reduction and classification of hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 11, pp. 7008–7022, Nov. 2014.

[4] L. Fang, S. Li, W. Duan, J. Ren, and J. A. Benediktsson, "Classification of hyperspectral images by exploiting spectral-spatial information of superpixel via multiple Kernels," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 12, pp. 6663–6674, Dec. 2015.

[5] G. Schohn and D. Cohn, "Less is more: Active learning with support vector machines," in *Proc. 17th ICML*, Jul. 2000, pp. 839–846.

[6] S. Tong and D. Koller, "Support vector machine active learning with applications to text classification," *J. Mach. Learn. Res.*, vol. 2, pp. 45–66, Nov. 2001.

[7] X. Ye, K. Sakai, L. O. Garciano, S.-I. Asada, and A. Sasao, "Estimation of citrus yield from airborne hyperspectral images using a neural network model," *Ecol. Model.*, vol. 198, nos. 3–4, pp. 426–432, Oct. 2006.

[8] E. Merényi, "Intelligent understanding of hyperspectral images through self-organizing neural maps," in *Proc. 2nd Int. Conf. CITSA*, Orlando, FL, USA, 2005, pp. 30–35.

[9] J. Li, J. Bioucas-Dias, and A. Plaza, "Spectral-spatial hyperspectral image segmentation using subspace multinomial logistic regression and Markov random fields," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 3, pp. 809–823, Mar. 2012.

[10] J. M. Bioucas-Dias, A. Plaza, G. Camps-Valls, P. Scheunders, N. M. Nasrabadi, and J. Chanussot, "Hyperspectral remote sensing data analysis and future challenges," *IEEE Geosci. Remote Sens. Mag.*, vol. 1, no. 2, pp. 6–36, Jun. 2013.

[11] J. M. Murphy and M. Maggioni, "Unsupervised clustering and active learning of hyperspectral images with nonlinear diffusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 3, pp. 1829–1845, Mar. 2019.
[12] A. Paoli, F. Melgani, and E. Pasolli, "Clustering of hyperspectral images based on multiobjective particle swarm optimization," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 12, pp. 4175–4188, Dec. 2009.
[13] Y. Chen, S. Ma, X. Chen, and P. Ghamisi, "Hyperspectral data clustering based on density analysis ensemble," *Remote Sens. Lett.*, vol. 8, no. 2, pp. 194–203, 2017.
[14] C. Cariou and K. Chehdi, "Unsupervised nearest neighbors clustering with application to hyperspectral images," *IEEE J. Sel. Topics Signal Process.*, vol. 9, no. 6, pp. 1105–1116, Sep. 2015.
[15] W. Zhu, V. Chayes, A. Tiard, S. Sanchez, D. Dahlberg, A. L. Bertozzi, S. Osher, D. Zosso, and D. Kuang, "Unsupervised classification in hyperspectral imagery with nonlocal total variation and primal-dual hybrid gradient algorithm," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 5, pp. 2786–2798, May 2017.
[16] E. Elhamifar and R. Vidal, "Sparse manifold clustering and embedding," in *Proc. Adv. Neural Inf. Process. Syst.*, pp. 55–63, 2011.
[17] E. Elhamifar and R. Vidal, "Sparse subspace clustering: Algorithm, theory, and applications," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 11, pp. 2765–2781, Nov. 2013.
[18] G. Papandreou, L.-C. Chen, K. P. Murphy, and A. L. Yuille, "Weakly- and semi-supervised learning of a deep convolutional network for semantic image segmentation," in *Proc. CVPR*, Feb. 2015, pp. 1742–1750.
[19] Y. Cui, G. Song, X. Wang, Z. Lu, and L. Wang, "Semisupervised classification of hyperspectral images based on tri-training algorithm with enhanced diversity," *J. Appl. Remote Sens.*, vol. 11, no. 4, Oct. 2017, Art. no. 045006.
[20] Z. Wang and J. Ye, "Querying discriminative and representative samples for batch mode active learning," *ACM Trans. Knowl. Discovery Data*, vol. 9, no. 3, Apr. 2015, Art. no. 17.
[21] X. Li and Y. Guo, "Adaptive active learning for image classification," in *Proc. CVPR*, Jun. 2013, pp. 859–866.
[22] M. Li and I. K. Sethi, "Confidence-based active learning," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 8, pp. 1251–1261, Aug. 2006.
[23] S. C. H. Hoi, R. Jin, J. Zhu, and M. R. Lyu, "Batch mode active learning and its application to medical image classification," in *Proc. ICML*, Jun. 2006, pp. 417–424.
[24] K. Brinker, "Incorporating diversity in active learning with support vector machines," in *Proc. Int. Conf. Mach. Learn.*, Washington, DC, USA, 2003, pp. 59–66.
[25] Z. Xu, K. Yu, V. Tresp, X. Xu, and J. Wang, "Representative sampling for text classification using support vector machines," in *Proc. 25th Eur. Conf. Inf. Retr. Res.*, 2003, pp. 393–407.
[26] E. Pasolli, F. Melgani, D. Tuia, F. Pacifici, and W. J. Emery, "SVM active learning approach for image classification using spatial information," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 4, pp. 2217–2233, Apr. 2014.
[27] R. Burbidge, J. J. Rowland, and R. D. King, "Active learning for regression based on query by committee," in *Intelligent Data Engineering and Automated Learning* (Lecture Notes in Computer Science), vol. 4881. Berlin, Germany: Springer-Verlag, 2007, pp. 209–218.
[28] Y. Freund, H. S. Seung, E. Shamir, and N. Tishby, "Selective sampling using the query by committee algorithm," *Mach. Learn.*, vol. 28, no. 2, pp. 133–168, 1997.
[29] I. Dagan and S. P. Engelson, "Committee-based sampling for training probabilistic classifiers," in *Proc. ICML*, San Francisco, CA, USA, 1995, pp. 150–157.
[30] D. Tuia, F. Ratle, F. Pacifici, M. F. Kanevski, and W. J. Emery, "Active learning methods for remote sensing image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 7, pp. 2218–2232, Jul. 2009.
[31] C. Li, Q. Gu, and Z. Cai, "Improved active learning algorithm and its application in hyperspectral classification," (in Chinese), *J. Huazhong Univ. Sci. Technol.*, vol. 41, no. 2, pp. 274–278, 2013.
[32] L. Copa, D. Tuia, M. Volpi, and M. Kanevski, "Unbiased query-by-bagging active learning for VHR image classification," *Proc. SPIE*, vol. 7830, Oct. 2010, Art. no. 78300K.
[33] S. Rajan, J. Ghosh, and M. M. Crawford, "An active learning approach to hyperspectral data classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 4, pp. 1231–1242, Apr. 2008.
[34] P. Mitra, B. U. Shankar, and S. K. Pal, "Segmentation of multispectral remote sensing images using active support vector machines," *Pattern Recognit. Lett.*, vol. 25, no. 9, pp. 1067–1074, Jul. 2004.
[35] B. Demir, C. Persello, and L. Bruzzone, "Batch-mode active-learning methods for the interactive classification of remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 3, pp. 1014–1031, Mar. 2011.
[36] Q. Shi, B. Du, and L. Zhang, "Spatial coherence-based batch-mode active learning for remote sensing image classification," *IEEE Trans. Image Process.*, vol. 24, no. 7, pp. 2037–2050, Jul. 2015.
[37] S. Patra and L. Bruzzone, "A novel SOM-SVM-based active learning technique for remote sensing image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 11, pp. 6899–6910, Nov. 2014.
[38] A. Rodriguez and A. Laio, "Clustering by fast search and find of density peaks," *Science*, vol. 344, no. 6191, pp. 1492–1496, Jun. 2014.
[39] B. J. Frey and D. Dueck, "Clustering by passing messages between data points," *Science*, vol. 315, no. 5814, pp. 972–976, Feb. 2007.
[40] S. Suzuki, M. Kakuta, T. Ishida, and Y. Akiyama, "Faster sequence homology searches by clustering subsequences," *Bioinformatics*, vol. 31, no. 8, pp. 1183–1190, 2015.
[41] J. C. Bezdek, "Pattern recognition with fuzzy objective function algorithms," *Adv. Appl. Pattern Recognit.*, vol. 22, no. 1171, pp. 203–239, 1981.
[42] X.-Q. Tang and P. Zhu, "Hierarchical clustering problems and analysis of fuzzy proximity relation on granular space," *IEEE Trans. Fuzzy Syst.*, vol. 21, no. 5, pp. 814–824, Oct. 2013.
[43] L. Zhang and J. You, "A spectral clustering based method for hyperspectral urban image," in *Proc. Joint Urban Remote Sens. Event*, Mar. 2017, pp. 1–3.
[44] C. Hou, F. Nie, D. Yi, and D. Tao, "Discriminative embedded clustering: A framework for grouping high-dimensional data," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 6, pp. 1287–1299, Jun. 2015.
[45] Y. Zhao, Y. Yuan, F. Nie, and Q. Wang, "Spectral clustering based on iterative optimization for large-scale and high-dimensional data," *Neurocomputing*, vol. 318, pp. 227–235, Nov. 2018.
[46] X. Zhang, L. Jiao, F. Liu, L. Bo, and M. Gong, "Spectral clustering ensemble applied to SAR image segmentation," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 7, pp. 2126–2136, Jul. 2008.
[47] G. Camps-Valls, T. V. B. Marsheva, and D. Zhou, "Semi-supervised graph-based hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 10, pp. 3044–3054, Oct. 2007.
[48] Y. Shao, N. Sang, C. Gao, and L. Ma, "Probabilistic class structure regularized sparse representation graph for semi-supervised hyperspectral image classification," *Pattern Recognit.*, vol. 63, pp. 102–114, Mar. 2017.
[49] B. Krishnapuram, L. Carin, M. A. T. Figueiredo, and A. J. Hartemink, "Sparse multinomial logistic regression: Fast algorithms and generalization bounds," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 6, pp. 957–968, Jun. 2005.
[50] Y. Xian and H. Hu, "Enhanced multi-dataset transfer learning method for unsupervised person re-identification using co-training strategy," *IET Comput. Vis.*, vol. 12, no. 8, pp. 1219–1227, Dec. 2018.
[51] M. A. Bencherif, Y. Bazi, A. Guessoum, N. Alajlan, F. Melgani, and H. AlHichri, "Fusion of extreme learning machine and graph-based optimization methods for active classification of remote sensing images," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 3, pp. 527–531, Mar. 2015.
[52] W. Liao, A. Pizurica, R. Bellens, S. Gautama, and W. Philips, "Generalized graph-based fusion of hyperspectral and LiDAR data using morphological features," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 3, pp. 552–556, Mar. 2015.
[53] F. Wang and C. Zhang, "Label propagation through linear neighborhoods," *IEEE Trans. Knowl. Data Eng.*, vol. 20, no. 1, pp. 55–67, Jan. 2008.
[54] L. Bruzzone, M. Chi, and M. Marconcini, "A novel transductive SVM for semisupervised classification of remote-sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 11, pp. 3363–3373, Nov. 2006.
[55] F. de Morsier, M. Borgeaud, V. Gass, J.-P. Thiran, and D. Tuia, "Kernel low-rank and sparse graph for unsupervised and semi-supervised classification of hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 6, pp. 3410–3420, Jun. 2016.
[56] Z. Wang, N. M. Nasrabadi, and T. S. Huang, "Semisupervised hyperspectral classification using task-driven dictionary learning with laplacian regularization," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 3, pp. 1161–1173, Mar. 2015.
[57] I. Dópido, J. Li, A. Plaza, and J. M. Bioucas-Dias, "Semi-supervised active learning for urban hyperspectral image classification," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Munich, Germany, Jul. 2012, pp. 1586–1589.
[58] W. Di and M. M. Crawford, "Active learning via multi-view and local proximity co-regularization for hyperspectral image classification," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 3, pp. 618–628, Jun. 2011.

[59] L. Wan, K. Tang, M. Li, Y. Zhong, and A. K. Qin, "Collaborative active and semisupervised learning for hyperspectral remote sensing image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 5, pp. 2384–2396, May 2015.
[60] L. Breiman, "Bagging predictors," *Mach. Learn.*, vol. 24, no. 2, pp. 123–140, 1996.
[61] R. Chen, Y. Cao, and H. Sun, "Multi-class image classification with active learning and semi-supervised learning," *Acta Autom. Sinica*, vol. 37, no. 8, pp. 954–962, 2011.
[62] S. Han, H. Zhang, and H. Zhou, "Decision tree classification algorithm based on relational degree function," (in Chinese), *J. Comput. Appl.*, vol. 2005, no. 11, pp. 2655–2657, 2005.
[63] A. Vlachos, "A stopping criterion for active learning," *Comput. Speech Lang.*, vol. 22, no. 3, pp. 295–312, 2008.
[64] R. Zhang and A. I. Rudnicky, "A large scale clustering scheme for kernel K-means," in *Proc. IEEE Int. Conf. Pattern Recog.*, Quebec, QC, Canada, Aug. 2002, pp. 289–292.
[65] Y. Cui, K. Xu, and Z. Lu, "Combination strategy of active learning for hyperspectral images classification," *J. Commun.*, vol. 39, no. 4, pp. 91–99, 2018.
[66] A. Samat, P. Du, M. H. A. Baig, S. Chakravarty, and L. Cheng, "Ensemble learning with multiple classifiers and polarimetric features for polarized SAR image classification," *Photogramm. Eng. Remote Sens.*, vol. 80, no. 3, pp. 239–251, Mar. 2014.
[67] R. O. Green, M. L. Eastwood, C. M. Sarture, T. G. Chrien, M. Aronsson, B. J. Chippendale, J. A. Faust, B. E. Pavri, C. J. Chovit, M. Solis, M. R. Olah, and O. Williams, "Imaging spectroscopy and the airborne visible/infrared imaging spectrometer (AVIRIS)," *Remote Sens. Environ.*, vol. 65, no. 3, pp. 227–248, Sep. 1998.
[68] L. Zhang, L. Zhang, D. Tao, and X. Huang, "Tensor discriminative locality alignment for hyperspectral image spectral-spatial feature extraction," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 1, pp. 242–256, Jan. 2013.
[69] J. D. Gibbons and S. Chakraborti, *Nonparametric Statistical Inference*. New York, NY, USA: Springer-Verlag, 2011.
[70] M. Belgiu and L. A. Drăguţ, "Random forest in remote sensing: A review of applications and future directions," *ISPRS J. Photogramm. Remote Sens.*, vol. 114, pp. 24-31, Apr. 2016.
[71] H. Cui, D. Huang, Y. Fang, L. Liu, and C. Huang, "Webshell detection based on random forest–gradient boosting decision tree algorithm," in *Proc. IEEE 3rd Int. Conf. Data Sci. Cyberspace (DSC)*, Jun. 2018, pp. 153–160.
[72] X. Meng. J. Bradley, B. Yavuz, E. Sparks, S. Venkataraman, D. Liu, J. Freeman, D. Tsai, M. Amde, S. Owen, D. Xin, R. Xin, M. J. Franklin, R. Zadeh, M. Zaharia, and A. Talwalkar, "Mllib: Machine learning in apache spark," *J. Mach. Learn. Res.*, vol. 17, no. 1, pp. 1235–1241, Jan. 2016.
[73] W. Fan and A. Bifet, "Mining big data: Current status, and forecast to the future," *ACM SIGKDD Explor. Newslett.*, vol. 14, no. 2, pp. 1–5, Dec. 2012.

YING CUI was born in Harbin, Hei longjiang, China, in 1979. She received the Ph.D. degree in signal and information processing and the Postdoctor's degree in computer science and technology from Harbin Engineering University, Harbin, China, in 2006 and 2009, respectively.

At the same time, she joined the Faculty, in 2005; and is currently an Associate Professor with the School of Information and Communication Engineering, HEU. Her research interests include intelligent signal processing, image processing, and wireless sensor network optimization. Dr. Cui is also a Reviewer for many journals, including the *Journal of Wuhan University*, the *Journal of Harbin Engineering University*, the *Journal of Applied Remote Sensing*, *Acta Optica Sinica*, and *Laser and Optoelectronics Progress*.

XIAOWEI JI received the B.E. degree in electronic information engineering from Northeast Forestry University (NEFU), Harbin, China, in 2017. She is currently pursuing the M.S. degree in information and communication engineering, Harbin engineering University (HEU). Her research interests include signal processing, remote sensing, and machine learning. She is currently participate in reviewing of IEEE Access.

HENG WANG received the master's degree in information and communication engineering from Harbin Engineering University, Harbin, China, in 2019. His research interests include hyperspectral image processing, computer science, and artificial intelligence.

KAI XU received the B.E. degree from the Harbin University of Science and Technology (HUST), China, in 2015, and the M.S. degree in information and communication engineering from Harbin Engineering University (HEU). His research interests include machine learning and image processing.

SHAOQIAO WU received the B.S. degree from Harbin Engineering University, China, in 2018, where he is currently pursuing the M.S degree. His research interests include the deep learning and remote sensing.

LIGUO WANG received the M.S. degree in natural science and the Ph.D. degree in engineering from the Harbin Institute of Technology (HIT), Harbin, China, and the Postdoctor's degree from Harbin Engineering University (HEU), China.

At the same time, he joined the Faculty, and is currently a Full Professor with the School of Information and Communication Engineering. Dr. Wang is currently a Senior Member of Chinese Institute of Electronics, Senior Member of China Institute of Communications (CIC), a Member of Chinese Society for Optical Engineering. He is also a Fellow of the International Society for Digital Earth (ISDE). He is also a Member of Chinese Society of Optical Engineering, Senior Committee of Photoelectric Technology of Astronautics Society. He is also a Reviewer for many journals, including the IEEE Transactions on Geoscience and Remote Sensing, the IEEE Geoscience and Remote Sensing Letters, and the *Journal of Electronics and Information Technology*.