

Local Binary Ensemble based Self-training for Semi-supervised Classification of Hyperspectral Remote Sensing Images

Pangambam Sendash Singh, Vijendra Pratap Singh, Manish Kumar Pandey, Subbiah Karthikeyan

Banaras Hindu University, Institute of Science, Department of Computer Science,
India

{pangambams.singh4, vijendrap.singh4, karthik}@bhu.ac.in, pandey.manish@live.com

Abstract. Supervised classification of hyperspectral remote sensing images is still challenging due to the scarcity of enough labelled samples. Semi-supervised methods have been adopted to handle this issue. Self-training is a popular semi-supervised technique which is widely used for training a classifier with limited labelled data and a large quantity of unlabeled data. However, traditional self-training approaches often give poor classification results in high dimensional data. In the current work, a novel efficient self-training approach for handling the deficiency of labelled samples for semi-supervised classification of hyperspectral remote sensing images is proposed. The proposed method first trains an ensemble of locally specialized supervised binary classifiers independently by using the dimensionally reduced spectral feature vectors of few available labelled samples. The trained local binary classifiers are then used to extend the labelled set by iterative addition of highly informative unlabeled samples to it by exploiting both the spectral and spatial information of the hyperspectral image. The classifiers are then retrained with the extended dataset in a batchwise manner and the procedure is repeated until adequate quantity of labelled samples are generated. Finally, a supervised multiclass classifier is trained on the extended dataset to produce the final classification map. Experimental results on two benchmark hyperspectral image datasets prove the effectiveness of the proposed method over supervised and traditional self-training based semi-supervised pixelwise classification approach in terms of different classification measures.

Keywords. Remote sensing, hyperspectral image analysis, machine learning, semi-supervised learning, self-training, ensembles.

1 Introduction

Remote sensing data has become the primary source of Geographical Information System (GIS) data. Data provided by the remote sensors to the GIS database are often multispectral or hyperspectral data which are in the form of images. Hyperspectral images (HSI) contain a large amount of spectral information which enables us to analyze an object or a scene very accurately. HSI classification, also termed as land cover classification in remote sensing community, has been widely applied in diverse areas such as target detection [7, 49], change detection [19, 21, 44], military defense, agriculture, water and forest resource management [1, 17, 25], disaster monitoring, etc.

With the advancement in both HSI data acquisition and machine learning technology, automated systems can be designed to perform HSI classification tasks. Supervised and unsupervised approaches are widely used in building such classification systems [36]. Supervised methods use the prior information of the classes to train a classification model. Traditional statistical models [33], support vector machines [4, 15], artificial neural networks [37], k -Nearest Neighbor algorithm [29], etc. are some of the popular algorithms for supervised HSI classification.

Performance of a supervised approach rely on the availability of a large number of labelled samples [3], which is not the case in remote sensing HSI data.

Though HSI can provide very rich spectral, spatial as well as temporal information, labeling each and every pixel requires proper ground survey of all the classes present over the area, which is a difficult and time-consuming process, while unlabeled samples are easily and abundantly available, which makes supervised approaches for land cover classification more challenging [39]. Semi-supervised methods which exploit both labelled as well as unlabeled data can be a solution to this problem [9] and the fact that obtaining a large amount of labelled samples is quite expensive as compared with unlabeled samples, has motivated many researchers to focus on semi-supervised methods [6, 43].

Generally, semi-supervised methods use automatic or semiautomatic labeling to provide the labels of the unlabeled samples. The unlabeled samples along with the assigned labels are added to the limited labelled set for retraining classifiers. This may sometimes lead to the presence of class label noise in the training set which may affect the efficiency of the classifiers directly. Moreover, HSI data usually consists of multiple classes which are often very much imbalanced in nature. Addressing both the problem of labelled sample deficiency and imbalanced data at the same time is of very crucial importance while building an automated HSI classification system.

In the current work, a novel efficient self-training approach for handling the deficiency of labelled training samples for semi-supervised HSI classification is proposed. First, an ensemble of locally specialized binary classifiers are trained on the limited labelled data by using spectral features through binary decomposition approach [2]. After that, the labelled set is iteratively extended by adding highly informative unlabeled samples to it. The quality as well as the label of an unlabeled sample is determined by exploiting both spatial and spectral information of the given HSI. The locally specialized binary classifiers are then retrained with the extended dataset in a batchwise manner. The whole procedure is repeated until adequate number of training samples are available. Finally, a supervised multiclass classifier is trained on the extended dataset for final HSI classification purpose.

The rest of the paper is organized as follows. Section 2 introduces related works about semi-supervised HSI classification. Section 3 introduces a brief review of self-training, binary decomposition of multiclass problems and clustering methods. The proposed method is explained in section 4. Section 5 gives the experimental setup. Results and discussions are presented in section 6. And finally, section 7 concludes the paper.

2 Related Work

There has been a number of works addressing on semi-supervised HSI classification methods.

In [6], the authors used a full family of composite kernels for a robust graph-based semi-supervised HSI classification. In [10], the authors considered the critical problem of non-convexity of the cost function optimization in semi-supervised SVMs for HSI classification by optimizing the cost function during the primal formulation rather than the dual formulation. In [32], a fuzzy c-means based iterative gathering of effective unlabeled samples was utilized for an ensemble based pixel-wise HSI classification. In [27], the authors used particle swarm optimization and fuzzy clustering to reduce the impact of incorrect labels and corrupted parameter values.

In [43], the authors proposed an efficient semi-supervised ensemble SVM that uses spectral similarity and mean shift based segmentation algorithm for dataset extension. In [39], compressive sensing technique was used for classification of multipectral satellite images with severe scarcity of labelled samples. In [42], the authors proposed an enhanced semi-supervised HSI classifier which is based on both neighbourhood information of the labelled and unlabeled samples and combination of two different classifiers. In [51], the authors used box-based smooth ordering and multiple 1D-embedding-based interpolation to address the problem of high dimensionality and the lack of labelled samples in HSI data.

In [31], the authors used weighted neighbourhood information and deep feature learning for labelling the unlabeled samples. In [30], the authors used two complementary regularizers that can preserve the local properties of both spectral

and spatial neighbourhood to improve graph based semi-supervised methods.

In [40], the authors designed an active learning protocol that aims at reducing the unlabeled sample search complexity to improve classification performance. In [38], the authors proposed a stable co-training approach, inspired by the Tracking-Learning-Detection, for classification of hyperspectral data by using both spatial and spectral features. In [18], the authors used generative adversarial networks to train on spectral-spatial features extracted from a HSI data cube by a three-dimensional bilateral filter (3DBF) for semi-supervised learning. In [46], the authors combined semi-supervised and active learning to mine both the representative and discriminative information by pseudo-labeling the unlabeled data with a supervised clustering technique.

In [28], the authors used weighted semi-supervised local discriminant analysis as the feature rotation tool to solve the problem of existing PCA based techniques that fail to take discriminative features during feature extraction. In [26], the authors proposed a semi-supervised convolutional neural network (CNN) with a ladder network that can automatically learn spectral-spatial features from complex HSI data cube. In [48], the authors used constrained Dirichlet process mixture model based clustering algorithm for labeling the unlabeled samples for dataset extension.

In [45], the authors used minimum trust evaluation and maximum uncertainty to estimate fusion evidence entropy of unlabeled samples during an iterative self-training based semi-supervised HSI classification framework. In [34], the authors used multi-grained scanning strategy to represent the full spectral and spatial relationships while building a deep learning based method called MugNet. In [13], the authors used extended label propagation and rolling guidance filtering methods for pseudo-labelling the unlabeled samples for semi-supervised training of a SVM model.

In [5], the authors used residual CNN (ResNet) and dual-strategy co-training for effective feature extraction and sample selection for a semi-supervised deep learning framework which is capable of reducing the dependence of deep learning methods on large-scale labeled

HSI data. In [22], the authors used PCA based edge-preserving features and extended morphological profiles to define a decision function on the basis of which the limited labeled set is extended on a large scale for HSI classification. In [3], a granular computing based self-training method was proposed for the semi-supervised classification of remote-sensing images.

In [50], the authors used multiple SVMs with different initial kernels to predict pseudo-labels independently. Consistency voting is applied to the resulting pseudo-labels for dataset augmentation. In [35], the authors proposed a novel semi-supervised spectral-spatial graph convolutional network that utilizes the adjacency nodes in the graph to add full spatial information embedded in the original HSI data. In [14], the authors combined ResNet with ensemble learning to extract preliminary image features and to establish discriminative image representations by exploring the intrinsic information of all available data for a semi-supervised scene classification for remote sensing images. In [52], the authors proposed a deep learning frame-work which combines textural features of gray level co-occurrence matrix into CNNs for HSI classification with limited labeled samples. A softmax neural network is employed for classification by using unsupervised textural features extracted by a PCA transformation and deep spectral features extracted by a CNN.

Most of the aforementioned methods paid attention to the extension of deficient labeled set by exploiting the spectral information without much consideration of the spatial information. The classification map of a given HSI, theoretically, depends on the spectral information only. However, due to the inherent limitations of HSI sensors, considering only the spectral information and ignoring the spatial information may result into the presence of class label noises that directly degrades the performance of classifiers. Exploitation in both spectral and spatial domain could enhance the quality of the unlabeled samples that have to be added to the labeled set as well as the performance of semi-supervised HSI classifiers.

To achieve this, a local binary ensemble based self-training method which exploits both the spectral and spatial information of the HSI to select high

quality, correct, informative and diverse unlabeled samples for semi-supervised classification of HSI is proposed in this current work.

3 Preliminaries

3.1 Self-Training

Self-training [8] is a popular semi-supervised approach which trains a classifier by using limited labeled samples and a huge pool of unlabeled samples. Let L and U be the sets of labeled and unlabeled samples respectively, with $|L| \ll |U|$. Let l be the number of classes and h be a supervised classifier. First, h is trained with the samples from L . The trained classifier h is then used to classify the samples from U . Then, a few most confident unlabeled samples from U along with the labels predicted by h are selected to include in L . Then, h is retrained with the updated labeled set L and the procedure is repeated. Finally, the supervised classifier h with updated labeled sample L is returned.

3.2 Multiclass to Binary-Class Decomposition

Classification is the process of mapping elements to a finite set of classes. In a multiclass classification problem, the number of classes is more than two. The increase in the number of classes often results into the increase in complexity and cost of a classifier. In such a case, decomposition of the multiclass classification problem into multiple binary classification problems that can be solved separately for only a subset of classes can be a solution [2, 24]. Decomposed binary classifiers return simpler decision boundaries that reduce the competence areas of each classifier, thus producing multiple local binary learners each dedicated to a binary sub-problem. A fusion of the results of these binary classifiers can be used to construct the classification result of the original problem [47].

3.3 Clustering

Clustering is the organization of a collection of objects into a finite number of homogeneous groups on the basis of some similarity measures in such a way that objects within a same group are more similar to each other than they are to objects in other groups [20].

3.3.1 k -means clustering

k -means Clustering [20] is a popular clustering technique in which a given collection of data is partitioned into k disjoint clusters. For an n number of data points $X = \{x_1, x_2, \dots, x_n\}$, k -means clustering algorithm works as follows:

1. Randomly initialize k cluster centers $X^* = \{x_1^*, x_2^*, \dots, x_k^*\}$.
2. For every data point $x_i \in X$ and for every cluster center $x_j^* \in X^*$, calculate the distance d_{ij} between x_i and x_j^* .
3. Assign x_i to cluster C_j if d_{ij} is minimum for every $1 \leq j \leq k$.
4. Update the positions of cluster centers by using Eq. (1) and go to step 2 until convergence:

$$x_j^* = \frac{1}{|C_j|} \sum_{x_i \in C_j} x_i, j = 1, 2, \dots, k. \quad (1)$$

3.3.2 Subtractive Clustering

Subtractive clustering [11] finds out the number of clusters and the locations of the initial cluster centers. For an n number of data points $X = \{x_1, x_2, \dots, x_n\}$, subtractive clustering algorithm works as follows:

1. For every data point $x_i \in X$, calculate initial potential P_i , which is given by:

$$P_i = \sum_{k=1}^n e^{-\frac{4\|x_k - x_i\|^2}{r_a^2}}, \quad (2)$$

where $r_a > 0$ is the hypersphere cluster radius which defines the radius of the neighborhoods.

2. Declare the data point (x^*) having the maximum potential (P^*) as the first cluster center.
3. Update the potential values of the data points by using the potential revision formula:

$$P_i = P_i - P^* e^{-\frac{4||x^* - x_i||^2}{r_b^2}} \quad (3)$$

where $r_b > 0$ is known as hypersphere penalty radius.

4. Declare the data point having the highest updated potential (P^*) as the next cluster center and go to step (3) until sufficient number of cluster centers are generated.

3.3.3 Hybrid Combination of Subtractive Clustering and k -means Clustering

Performance of a k -means clustering algorithm depends on the locations of initial cluster centers, which are randomly initialized in a traditional k -means clustering algorithm [23]. Also, a priori knowledge of the value of k is also must. A hybrid combination of subtractive clustering with k -means clustering algorithm can improve the performance of k -means clustering algorithm. In the hybrid combination, subtractive clustering can be used to find out the number of clusters (k) and the initial locations of the cluster centers $x_1^*, x_2^*, \dots, x_k^*$ for a better k -means clustering algorithm.

4 Proposed Methodology

Let L and U be the set of labeled and unlabeled HSI samples. Let d be the number of spectral bands present in the HSI and $Y = \{l_1, l_2, \dots, l_l\}$ be the set of class labels with $|Y| = l$.

Theoretically, the class labels depend on the spectral information only, however it is not enough to decide the label based on the spectral information only due to various reasons [30]. More exploitation of both spectral and spatial information could give better classification results. Following are the main steps of the proposed methodology:

1. Dimensionality reduction along spectral domain by data fusion.

2. Supervised learning on limited labeled set by using locally specialized binary classifiers through binary decomposition.

3. Self-training until adequate samples are generated:

- (a) Iteratively select high quality, informative unlabeled samples to extend the limited labeled dataset through the exploitation of local spatial and global spectral features.

- (b) Retrain the locally specialized binary classifiers on the extended dataset in a batchwise manner.

4. Train an efficient multiclass supervised classifier with the extended labeled dataset to produce final HSI classification map.

Fig.1 shows the block diagram of the steps involved in the proposed self-training based semi-supervised HSI classification. In the following sections, each step is discussed in detail.

4.1 Dimensional Reduction Along Spectral Domain by Data Fusion

Band averaging method is used for dimensional reduction of the HSI along the spectral domain. In this method, a given HSI having d -spectral bands is spectrally partitioned into m sub-groups of hyperspectral data, each having d/m adjacent spectral bands. After that, the average band is calculated for each subgroup so as to obtain a dimensionally reduced hyperspectral data having $m(< d)$ bands. The advantage of this method over other transform-based methods like PCA, ICA, etc. is that pixel values of the reduced data are still somehow related to the reflectance value of the original HSI.

4.2 Supervised Learning on Limited Labeled Set by using Locally Specialized Binary Classifiers Through Binary Decomposition

Binary decomposition of the multiclass HSI classification problem (explained in section 3.2) is achieved by creating l binary classification problems, one for each class [2]. That is, for each

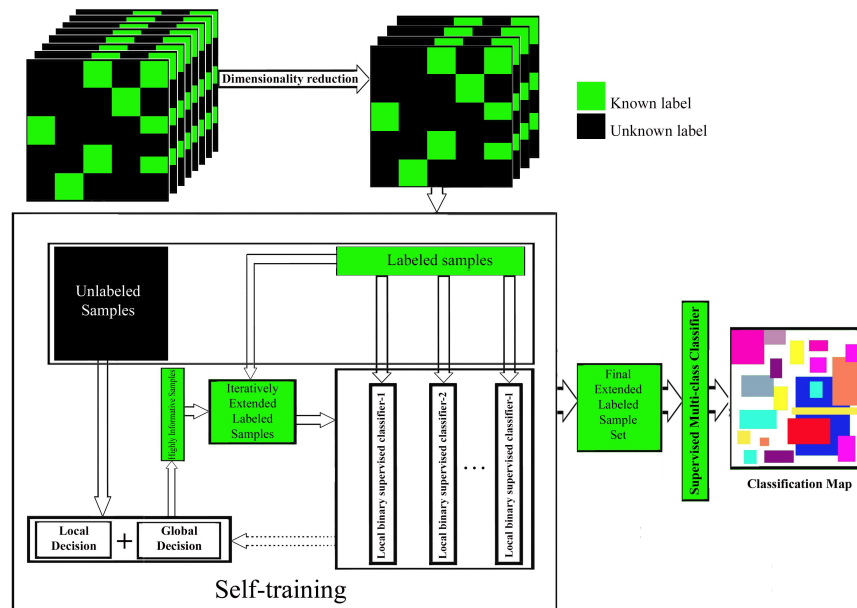


Fig. 1. Block diagram of the proposed self-training based semi-supervised method for HSI classification

class $l_i \in Y$, a binary classifier h_i is trained in such a way that samples labeled $y = l_i$ are considered as positive class and all other samples considered as negative class, thus creating l number of independent local binary learners $\{h_i\}, i = 1, \dots, l$, each dedicated to a specific binary sub-problem. Proper rebalancing is done so that these binary datasets are more or less balanced.

4.3 Self-Training: Selection of High-Quality Unlabeled Training Samples for Self-Training

The success of a self-training based semi-supervised technique depends on the selection of informative and diverse unlabeled samples along with correct class labels. In our approach the quality as well as the class label of an unlabeled sample is determined on the basis of global spectral as well as local spatial information of the given HSI.

4.3.1 Global Spectral Decision of an Unlabeled Sample

Samples belonging to each class are clustered independently into some finite number of clusters by using hybrid clustering combination (explained in section 3.3.3). Global spectral decision of an unlabeled sample u_i is taken on the basis of spectral Euclidean distance between u_i and cluster centers of each class. The detail steps are listed in the following:

1. For each class C_j , find spectral cluster centers $\{x_1^{j*}, x_2^{j*}, \dots, x_{n_j}^{j*}\}$ by using the hybrid clustering algorithm.
2. Find the spectral distance d_{ij} between u_i and each class C_j .

$$d_{ij} = \min_{k=1}^{n_j} \left\{ \left\| u_i - x_k^{j*} \right\| \right\}, \quad (4)$$

3. Find the class j having the minimum spectral distance, i.e. the value of j that satisfies the inequality (5).

$$d_{ij} \leq d_{ik} \forall k = 1, 2, \dots, l, \quad (5)$$

4. Assign j as the final conclusion of the global decision if the local binary classifier h_j also classifies u_i as a sample belonging C_j .

$$h_j(u_i) = 1 \Rightarrow \chi_{glo}^i = j. \quad (6)$$

4.3.2 Local Spatial Decision of an Unlabeled Sample

This method is based on the fact that the neighboring pixels of a HSI in a homogeneous region usually belong to a same class. This method can reduce the labeling error of an unlabeled pixel by exploiting the local spatial neighborhood information. The detail steps are as follows:

1. By taking u_i as center, find its s -square neighborhoods $\{N_1, N_2, \dots, N_s\}$.
2. For each class C_j , find the weighted score of u_i that relates the probability of u_i belonging to C_j by using the formula:

$$S_{ij} = \sum_{N_k \in C_j} \frac{1}{\|u_i - N_k\|}. \quad (7)$$

Inverse of Euclidean distance is used because nearby labeled samples should have more decision power than the ones that are far away from u_i .

3. The class having the maximum score value is assigned as the final conclusion of the local decision of u_i .

$$\chi_{loc}^i = j : S_{ij} \geq S_{ik} \forall k = 1, 2, \dots, l. \quad (8)$$

4.3.3 Final Decision and Retraining

Final decision is made on the basis of both local and global decisions. For an unlabeled sample $u_i \in U$, if the local decision agrees with the global decision, then u_i is considered as a high-quality sample, so u_i along with its predicted label has to be added to the labeled set L for dataset extension:

$$\chi_{glo}^i = \chi_{loc}^i \Leftrightarrow L = L \cup \{u_i\}. \quad (9)$$

After the iterative updation of the labeled set, the locally specialized binary classifiers are retrained

on the extended dataset and the spectral cluster centers are also updated accordingly in a batch-wise fashion. The whole procedure of extending the labeled sample set and retraining the classifiers on the extended dataset is repeated until adequate quantity of labeled samples are generated.

After generation of an adequate number of labeled samples, a supervised multiclass classifier is trained on the extended dataset to produce final image classification map.

5 Experimental Setup

5.1 Datasets

Two benchmark HSI datasets with different spectral and spatial resolutions are used to evaluate the performance of the proposed approach in real scenario.

The first image is the University of Pavia dataset which was acquired with the Reflective Optics System Imaging Spectrometer (ROSIS) optical sensor, with spatial resolution of 1.3 m per pixel and spectral coverage ranging from 0.43 to 0.86 μm , over an urban area surrounding the University of Pavia, Italy. The image has 115 bands of size 610×340 pixels out of which 12 noisy bands were removed.

The ground truth data contains nine classes of interest viz. trees, asphalt, bitumen, gravel, metal sheets, shadow, bricks, meadows, and bare soil.

The other image is the Indian Pines dataset which was acquired by the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) sensor, with spatial resolution of 20 m per pixel and spectral coverage ranging from 0.4 to 2.5 μm , over the agricultural Indian Pine test site of North-western Indiana. The image has 224 spectral bands of size 145×145 pixels out of which twenty water adsorption bands were removed. The ground truth data contains sixteen classes of interest consisting of agricultural area, forest and other natural perennial vegetation.

The datasets are available in the website¹. The sample band of both the datasets and their

¹<http://lesun.weebly.com/hyperspectral-data-set.html>

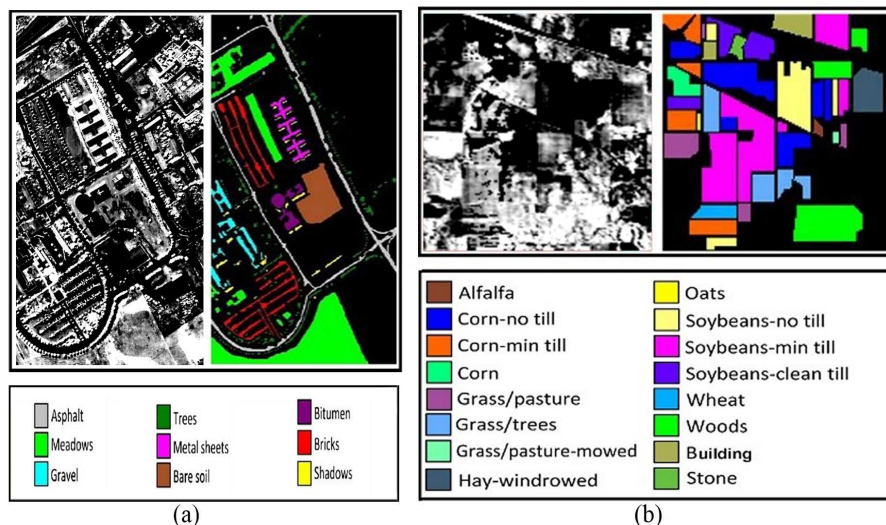


Fig. 2. Sample band and corresponding ground truth of (a) the University of Pavia and (b) the Indian Pines hyperspectral image data

corresponding ground truths are presented in Fig. 2.

5.2 Experimental Strategy

The effectiveness of the proposed semi-supervised algorithm for HSI classification is accessed by using 10-fold cross validation method. In each fold, 90% of the dataset is used as training set and the remaining 10% as the testing set.

Meanwhile, the training set is further subdivided into labeled and unlabeled set. The proposed method is evaluated twice in each dataset with different labeled unlabeled sample ratios. In the first round, in each fold, 40% of the training set is treated as labeled data and the remaining 60% as unlabeled data, and in the second round, 30% as labeled data and remaining 70% as unlabeled data. The performance of the proposed self-training method is compared with supervised and traditional self-training based semi-supervised HSI classification methods. Binary support vector machines [12] are used for building local binary classifiers due to its superior result in terms of HSI classification accuracy and robustness to high dimensional data [16, 37, 41]. k -nearest neighbor classification algorithm is used for supervised

training using the extended training set for final pixelwise image classification purpose.

5.3 Quality Indexes

Four quality indexes namely average classification accuracy (AA), precision (P), recall (R) and F_1 -score (F_1) are adopted to evaluate the performance of the proposed approach. For an l -class multiclass classification problem, these measures are given by:

$$AA = \frac{\sum_{i=1}^l \frac{tp_i + tn_i}{tp_i + fn_i + fp_i + tn_i}}{l}, \quad (10)$$

$$P = \frac{\sum_{i=1}^l \frac{tp_i}{tp_i + fp_i}}{l}, \quad (11)$$

$$R = \frac{\sum_{i=1}^l \frac{tp_i}{tp_i + fn_i}}{l}, \quad (12)$$

$$F_1 = 2 \times \frac{P \times R}{P + R}, \quad (13)$$

where tp_i , fp_i , fn_i and tn_i are true positive, false positive, false negative and true negative counts respectively for an arbitrary class C_i . Note that

Table 1. University of Pavia HSI dataset classification accuracy comparison under two different labeled unlabeled sample ratios

Class	L:U = 40:60			L:U = 30:70		
	S	SS	LBESS	S	SS	LBESS
Asphalt	75.358	76.377	93.122	75.4573	76.647	92.856
Meadows	77.749	77.504	88.079	76.897	76.423	87.586
Gravel	11.231	17.642	69.884	12.742	21.951	69.211
Trees	84.806	94.652	90.983	87.118	97.584	90.617
Metal sheets	99.252	99.329	99.632	99.026	99.179	99.486
Bare Soil	88.245	94.072	96.671	93.572	94.901	95.101
Bitumen	44.412	65.412	75.724	39.121	64.741	72.988
Bricks	60.739	60.786	79.935	59.207	59.725	79.977
Shadows	91.645	93.451	98.891	90.214	93.457	98.635
Quality indexes						
Avg. Accuracy	70.381	75.469	88.102	70.372	76.067	87.384
Precision	0.626	0.626	0.848	0.614	0.614	0.844
Recall	0.761	0.779	0.853	0.766	0.781	0.843
F ₁ -Measure	0.603	0.607	0.848	0.591	0.594	0.842

Table 2. Indian Pines HSI dataset classification accuracy comparison under two different labeled unlabeled sample ratios

Class	L:U = 40:60			L:U = 30:70		
	S	SS	LBESS	S	SS	LBESS
Alfalfa	11.196	17.635	84.452	9.194	11.648	88.333
Corn_N	65.366	60.245	74.414	36.756	5.675	75.432
Corn_M	13.423	12.174	72.266	12.142	13.637	75.115
Corn	12.123	15.637	59.174	3.219	11.362	64.943
Grass_P	25.679	20.768	82.595	20.622	31.472	81.493
Grass_T	62.399	52.166	85.538	60.537	51.381	83.772
Grass_PM	11.073	21.894	90.166	9.582	17.674	90.833
Hay_W	80.376	81.078	95.999	81.188	80.824	93.176
Oats	5.547	11.914	61.666	7.754	21.741	70.833
Soybean_N	11.634	10.161	73.966	11.719	34.623	71.998
Soybean_M	40.776	41.721	80.485	37.531	37.724	79.589
Soybean_C	9.754	17.612	66.714	10.214	44.213	65.079
Wheat	16.765	15.411	89.838	14.267	27.124	87.639
Woods	72.812	72.732	90.902	72.193	72.708	90.487
Buildings	22.219	27.214	64.081	23.465	40.127	59.721
Stone	92.196	93.141	98.092	92.147	94.157	98.092
Quality indexes						
Avg. Accuracy	34.583	35.718	79.396	31.408	37.255	79.783
Precision	0.372	0.368	0.766	0.353	0.352	0.751
Recall	0.531	0.552	0.763	0.527	0.551	0.757
F ₁ - Measure	0.355	0.346	0.761	0.329	0.318	0.749

a measure is calculated by averaging the same measures calculated for C_1, \dots, C_l in each fold.

6 Results and Discussions

The experimental results with University of Pavia and Indian Pines HSI datasets in terms of different classification measures under two different scenarios are listed in Table 1 and Table 2 respectively.

Supervised training, traditional self-training based semi-supervised training and the proposed local binary ensemble based self-training semi-supervised method are abbreviated as S, SS and LBESS respectively.

The comparative analysis shows that the proposed method outperforms supervised learning and traditional self-training based semi-supervised learning for HSI classification under the scarcity of labeled samples.

Significant improvement can be seen in all the quality measures in the proposed method. Better values in precision, recall and F_1 -Measure imply the low misclassification errors. From these results, it can also be concluded that the proposed approach selects highly informative, diverse unlabeled samples for self-training purpose and assigns correct class labels efficiently.

The use of simple binary classifiers while building the local ensembles makes the proposed method computationally less expensive. Combination of local decision based on spatial information and global decision based on spectral information along with the classification results of local binary classifiers while selecting the unlabeled samples ensures the selection of high-quality informative samples along with correct class label for dataset extension. This proposed approach can also be used to solve the problem of data imbalance effectively.

7 Conclusion and Future Work

A local binary ensemble based self-training method for semi-supervised HSI classification has been proposed in the current work. The proposed wrapper method iteratively extends the limited labeled set by selecting high-quality, informative and diverse unlabeled samples through the exploitation of both spectral and spatial information of the HSI.

Binary SVMs were used while building local binary ensembles for self-training and k -nearest neighbor classifier was used for supervised training on the extended dataset to produce final image classification map.

Global spectral based and local spatial based decisions were utilized to decide the class label of an unlabeled sample. A hybrid clustering method along with classification results given by the local binary classifiers was used for taking global decisions and a measure which relates to weighted Euclidean distance between the unlabeled sample and nearby labeled samples was used for taking local decisions.

Experimental results on two benchmark HSI datasets show that the proposed method efficiently outperforms purely supervised learning and traditional self-training based semi-supervised learning for HSI classification when the labeled samples are deficient. The proposed method can also be used to solve the problem of data imbalance very effectively. Identification of better feature extraction techniques for dimensional reduction, optimization of the classifiers and decision parameters will be our future research.

Acknowledgements

First author would like to thank and acknowledge the University Grants Commission (UGC), New Delhi for providing fellowship to pursue his research through UGC-Junior Research Fellowship Scheme. Further, the authors would like to thank the Department of Science and Technology (DST), New Delhi for the technical support provided through DST-PURSE scheme.

References

1. **Adam, E., Mutanga, O., & Rugege, D. (2010).** Multispectral and hyperspectral remote sensing for identification and mapping of wetland vegetation: A review. *Wetlands Ecology and Management*, Vol. 18, No. 3, pp. 281–296.
2. **Allwein, E. L., Schapire, R. E., & Singer, Y. (2001).** Reducing multiclass to binary: A unifying approach for margin classifiers. *Journal of Machine Learning Research*, Vol. 1, pp. 113–141.
3. **Aydav, P. S. S. & Minz, S. (2020).** Granulation-based self-training for the semi-supervised classification of remote-sensing images. *Granular Computing*, Vol. 5, pp. 309–327.
4. **Bazi, Y. & Melgani, F. (2006).** Toward an optimal svm classification system for hyperspectral remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 44, No. 11, pp. 3374–3385.
5. **Bei Fang, Ying Li, Haokui Zhang, J. C.-W. C. (2018).** Semi-supervised deep learning classification for hyperspectral image based on dual-strategy sample selection. *Remote Sensing*, Vol. 10, No. 4.
6. **Camps-Valls, G., Bandos Marsheva, T. V., & Zhou, D. (2007).** Semi-supervised graph-based hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 45, No. 10, pp. 3044–3054.
7. **Cao, Y., Zhang, J., Tian, Q., Zhuo, L., & Zhou, Q. (2015).** Salient target detection in hyperspectral images using spectral saliency. *IEEE China Summit and International Conference on Signal and Information Processing (ChinaSIP)*, pp. 1086–1090.
8. **Chapelle, O., Scholkopf, B., & Zien, A. (2010).** *Semi-Supervised Learning*. The MIT Press, 1st edition.
9. **Chapelle, O., Sindhwani, V., & Keerthi, S. S. (2008).** Optimization techniques for semi-supervised support vector machines. *Journal of Machine Learning Research*, Vol. 9, pp. 203–233.
10. **Chi, M. & Bruzzone, L. (2007).** Semisupervised classification of hyperspectral images by SVMs optimized in the primal. *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 45, No. 6, pp. 1870–1880.
11. **Chiu, S. (1994).** Fuzzy model identification based on cluster estimation. *Journal of Intelligent and Fuzzy Systems*, Vol. 2, No. 3, pp. 267–278.
12. **Cortes, C. & Vapnik, V. (1995).** Support vector networks. *Machine Learning*, Vol. 20, No. 3, pp. 273–297.
13. **Cui, B., Xie, X., Hao, S., Cui, J., & Lu, Y. (2018).** Semi-supervised classification of hyperspectral images based on extended label propagation and rolling guidance filtering. *Remote Sensing*, Vol. 10, No. 4, pp. 515.
14. **Dai, X., Wu, X., Wang, B., & Zhang, L. (2019).** Semisupervised scene classification for remote sensing images: A method based on convolutional neural networks and ensemble learning. *IEEE Geoscience and Remote Sensing Letters*, Vol. PP, pp. 1–5.
15. **Demir, B. & Erturk, S. (2007).** Hyperspectral image classification using relevance vector machines. *IEEE Geoscience and Remote Sensing Letters*, Vol. 4, No. 4, pp. 586–590.
16. **Foody, G. M. & Mathur, A. (2004).** A relative evaluation of multiclass image classification by support vector machines. *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 42, No. 6, pp. 1335–1343.
17. **Govender, M., Chetty, K., & Bulcock, H. (2009).** A review of hyperspectral remote sensing and its application in vegetation and water resource studies. *Water SA*, Vol. 33, No. 2, pp. 145–151.
18. **He, Z., Liu, H., Wang, Y., & Hu, J. (2017).** Generative adversarial networks-based semi-supervised learning for hyperspectral image classification. *Remote Sensing*, Vol. 9, No. 10, pp. 1042.
19. **Huang, F., Yu, Y., & Feng, T. (2018).** Hyperspectral remote sensing image change detection based on tensor and deep learning. *Journal of Visual Communication and Image Representation*, Vol. 58, pp. 233–244.
20. **Jain, A. K., Murty, M. N., & Flynn, P. J. (1999).** Data clustering: A review. *ACM Comput. Surv.*, Vol. 31, No. 3, pp. 399–404.
21. **Jamali, S., Jönsson, P., Eklundh, L., Ardö, J., & Seaquist, J. (2015).** Detecting changes in vegetation trends using time series segmentation. *Remote Sensing of Environment*, Vol. 156, pp. 182–195.
22. **Kang, X., Zhuo, B., & Duan, P. (2019).** Semi-supervised deep learning for hyperspectral image classification. *Remote Sensing Letters*, Vol. 10, No. 4, pp. 353–362.
23. **Khan, S. S. & Ahmad, A. (2004).** Cluster center initialization algorithm for K-means clustering.

- Pattern Recognition Letters*, Vol. 25, No. 11, pp. 1293–1302.
24. **Krawczyk, B., Woźniak, M., & Herrera, F. (2015).** On the usefulness of one-class classifier ensembles for decomposition of multi-class problems. *Pattern Recognition*, Vol. 48, No. 12, pp. 3969–3982.
 25. **Lanthier, Y., Bannari, A., Haboudane, D., Miller, J. R., & Tremblay, N. (2008).** Hyperspectral data segmentation and classification in precision agriculture: A multi-scale analysis. *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, volume 2, IEEE, pp. II-585–II-588.
 26. **Liu, B., Yu, X., Zhang, P., Tan, X., Yu, A., & Xue, Z. (2017).** A semi-supervised convolutional neural network for hyperspectral image classification. *Remote Sensing Letters*, Vol. 8, No. 9, pp. 839–848.
 27. **Liu, Y., Zhang, B., Wang, L.-m., & Wang, N. (2013).** A self-trained semisupervised SVM approach to the remote sensing land cover classification. *Computers & Geosciences*, Vol. 59, pp. 98–107.
 28. **Lu, X., Zhang, J., Li, T., & Zhang, Y. (2017).** Hyperspectral image classification based on semi-supervised rotation forest. *Remote Sensing*, Vol. 9, No. 9.
 29. **Ma, L., Crawford, M. M., & Tian, J. (2010).** Local manifold learning-based k -nearest-neighbor for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 48, No. 11, pp. 4099–4109.
 30. **Ma, L., Ma, A., Ju, C., & Li, X. (2016).** Graph-based semi-supervised learning for spectral-spatial hyperspectral image classification. *Pattern Recognition Letters*, Vol. 83, pp. 133–142.
 31. **Ma, X., Wang, H., & Wang, J. (2016).** Semisupervised classification for hyperspectral image based on multi-decision labeling and deep feature learning. *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 120, pp. 99–107.
 32. **Maulik, U. & Chakraborty, D. (2011).** A self-trained ensemble with semisupervised SVM: An application to pixel classification of remote sensing imagery. *Pattern Recognition*, Vol. 44, No. 3, pp. 615–623.
 33. **Mohamed, R. & Farag, A. (2005).** Advanced algorithms for bayesian classification in high dimensional spaces with applications in hyperspectral image segmentation. *IEEE International Conference on Image Processing 2005*, volume 2, IEEE, pp. II-646.
 34. **Pan, B., Shi, Z., & Xu, X. (2018).** MugNet: Deep learning for hyperspectral image classification using limited samples. *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 145, pp. 108–119.
 35. **Qin, A., Shang, Z., Tian, J., Wang, Y., Zhang, T., & Tang, Y. Y. (2019).** Spectral–spatial graph convolutional networks for semisupervised hyperspectral image classification. *IEEE Geoscience and Remote Sensing Letters*, Vol. 16, No. 2, pp. 241–245.
 36. **Richards, J. A. (2013).** *Remote Sensing Digital Image Analysis*, volume 5. Springer Berlin Heidelberg, Berlin, Heidelberg.
 37. **Roli, F., Serpico, S., & Bruzzone, L. (1996).** Classification of multisensor remote-sensing images by multiple structured neural networks. *Proceedings of 13th International Conference on Pattern Recognition*, volume 4, IEEE, pp. 180–184 vol.4.
 38. **Romaszewski, M., Głomb, P., & Cholewa, M. (2016).** Semi-supervised hyperspectral classification from a small number of training samples using a co-training approach. *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 121, pp. 60–76.
 39. **Roy, M., Melgani, F., Ghosh, A., Blanzieri, E., & Ghosh, S. (2015).** Land-cover classification of remotely sensed images using compressive sensing having severe scarcity of labeled patterns. *IEEE Geoscience and Remote Sensing Letters*, Vol. 12, No. 6, pp. 1257–1261.
 40. **Samat, A., Li, J., Liu, S., Du, P., Miao, Z., & Luo, J. (2016).** Improved hyperspectral image classification by active learning using pre-designed mixed pixels. *Pattern Recognition*, Vol. 51, pp. 43–58.
 41. **Shao, Y. & Lunetta, R. S. (2012).** Comparison of support vector machine, neural network, and CART algorithms for the land-cover classification using limited training data points. *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 70, pp. 78–87.
 42. **Tan, K., Hu, J., Li, J., & Du, P. (2015).** A novel semi-supervised hyperspectral image classification approach based on spatial neighborhood information and classifier combination. *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 105, pp. 19–29.
 43. **Tan, K., Li, E., Du, Q., & Du, P. (2014).** An efficient semi-supervised classification approach for hyperspectral imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 97, pp. 36–45.

44. **Thyagarajan, K. K. & Vignesh, T. (2019)**. Soft computing techniques for land use and land cover monitoring with multispectral remote sensing images: A review. *Archives of Computational Methods in Engineering*, Vol. 26, No. 2, pp. 275–301.
45. **Wang, C., Xu, Z., Wang, S., & Zhang, H. (2018)**. Semi-supervised classification framework of hyperspectral images based on the fusion evidence entropy. *Multimedia Tools and Applications*, Vol. 77, No. 9, pp. 10615–10633.
46. **Wang, Z., Du, B., Zhang, L., Zhang, L., & Jia, X. (2017)**. A novel semisupervised active-learning algorithm for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 55, No. 6, pp. 3071–3083.
47. **Woźniak, M., Graña, M., & Corchado, E. (2014)**. A survey of multiple classifier systems as hybrid systems. *Information Fusion*, Vol. 16, No. 1, pp. 3–17.
48. **Wu, H. & Prasad, S. (2018)**. Semi-supervised deep learning using pseudo labels for hyperspectral image classification. *IEEE Transactions on Image Processing*, Vol. 27, No. 3, pp. 1259–1270.
49. **Xu, Y., Du, Q., & Younan, N. H. (2017)**. Particle swarm optimization-based band selection for hyperspectral target detection. *IEEE Geoscience and Remote Sensing Letters*, Vol. 14, No. 4, pp. 554–558.
50. **Yang, S., Hou, J., Jia, Y., Mei, S., & Du, Q. (2019)**. Pseudolabel guided kernel learning for hyperspectral image classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, Vol. 12, No. 3, pp. 1000–1011.
51. **Ye, Z., Li, H., Song, Y., Wang, J., & Benediktsson, J. A. (2016)**. A novel semi-supervised learning framework for hyperspectral image classification. *International Journal of Wavelets, Multiresolution and Information Processing*, Vol. 14, No. 02, pp. 1640005.
52. **Zhao, W., Li, S., Li, A., Zhang, B., & Li, Y. (2019)**. Hyperspectral images classification with convolutional neural network and textural feature using limited training samples. *Remote Sensing Letters*, Vol. 10, No. 5, pp. 449–458.

Article received on 29/10/2019; accepted on 06/03/2020.
Corresponding author is Pangambam Sendash Singh.