

A SPECTRAL-SPATIAL AUGMENTED ACTIVE LEARNING METHOD FOR HYPERSPPECTRAL IMAGE CLASSIFICATION

Sh.Falahatnejad*, A.Karami

Physics Faculty, Shahid Bahonar University of Kerman, Kerman, Iran
sh.falahat@phy.uk.ac.ir, akarami@uk.ac.ir

Commission IV, WG IV/3

KEY WORDS: Hyperspectral Image Classification , Active Learning , Kernel Minimum Noise Fraction , Shearlet Transform , Data Augmentation , Outlier Elimination , Kolmogorov-Smirnov Test

ABSTRACT:

In this paper, a new classification technique for hyperspectral images (HSIs) based on an augmented active learning (AL) is introduced. The proposed method consists of two main steps: first, a 2-D non-subsampled shearlet transform (NSST) is applied to each spectral band of HSIs to extract the spatial features. After that, the kernel minimum noise fraction (KMNF) is used to reduce the spectral dimension. Second, the classification task using an augmented active learning technique is performed. For this purpose, an iterative process is considered. At each iteration, a discriminative sample selection and augmentation are used to create the training set. Then, the support vector machine (SVM) is iteratively applied to the training set. In the proposed method, the most informative samples are selected by a new query function combination of a posterior probability-based uncertainty and angle-based diversity criteria. The augmentation strategy during the training process is chosen by two-sample Kolmogorov-Smirnov test and the existing outliers are removed by k-means clustering. Finally, the proposed algorithm is applied to the real datasets and compared with three state-of-the-art AL algorithms. The obtained results show that the proposed method significantly increases accuracy considering the most informative samples.

1. INTRODUCTION

Recently, hyperspectral image classification especially supervised approaches has gained particular attention in many practical applications such as agriculture, mineralogy, environmental studies, etc. (Camps-Valls et al., 2014). The accuracy of supervised techniques is highly dependent on the quality of annotated dataset which is provided by user. Preparing enough labeled samples is really time-consuming and an expensive process (He et al., 2017).

Active learning (AL) approaches could significantly help to increase the classification accuracy by detecting the most informative pixels during the training and decrease the labeling effort. In fact, AL selects the most informative unlabeled samples from a data pool to refine the learning performance (Persello and Bruzzone, 2014). In recent years, AL has been extensively studied in the field of hyperspectral image classification. At first, a semi-supervised multinomial logistic regression model combined with entropy (EP)-based active selection strategy (Li et al., 2010) was presented. Then, the AL strategies combined with Bayesian classification method and loopy belief propagation technique (Li et al., 2011b, Li et al., 2011a) were investigated. After that, an AL framework based on Markov random field (MRF) (Sun et al., 2015) was introduced. In addition, some works based on the combination of the AL strategy and deep learning have been studied for HSI classification (Li, 2015, Liu et al., 2016, Haut et al., 2018). Particularly, a technique that integrates a multiclass level uncertainty (MCLU) active learning criterion with a stacked autoencoder (SAE)-based neural network (Li, 2015) was designed. In (Liu et al., 2016) a strategy to join the restricted Boltzmann machine (RBM) with a weighted incremental dictionary learning criterion was proposed. A method that utilizes six AL sampling criteria, such as

maximum EP, breaking ties (BT), random acquisition, mutual information (MI), etc., with Bayesian-convolutional neural network (BCNN) was presented (Haut et al., 2018). In (Paoletti et al., 2020), the performance of capsule networks (CapsNets) was enhanced and led to better results of HSI classification by utilizing a new AL method based on BT criterion. Additionally, a decoupled network with an active learning strategy (DCN-AL) is introduced (Bai et al., 2020). This technique considers both intra-class and inter-class variations and extracts features more efficiently.

Based on the number of informative samples which are elected at each iteration, two kinds of strategies for AL algorithms are considered: the single and batch modes. At each iteration, single mode selects the single most informative sample while the batch mode selects a batch of the most informative ones. The batch mode strategy usually achieves higher classification accuracy (Tuia et al., 2009).

Recently, two main criteria called uncertainty and diversity have been introduced. They are integrated to select the most informative samples (Patra and Bruzzone, 2011).

The uncertainty criterion selection aims at selecting a batch of unlabeled samples that have the lowest classification confidence. The uncertainty based batch mode approaches are divided into three heuristic categories (Tuia et al., 2009). First, posterior probability-based techniques such as best versus second best (BvSB) (Li and Zhang, 2016), BT (Ahmad et al., 2019) and Kullback-Leibler (KL)-Max strategy (Jun and Ghosh, 2008, Rajan et al., 2008) which calculate the class confidence to evaluate classification uncertainty of each unlabeled sample (Yu et al., 2015). A well-known AL method in this category that considers both spectral and spatial information using maximum a posterior marginal (MPM) and loopy belief propagation (LB) then utilizes BT to select informative samples is MPM-LBP-BT (Li et

* Corresponding author

al., 2013). Second, large margin based approaches for example margin sampling (MS) (Tuija et al., 2009) and MCLU (Demir et al., 2011) which relies on SVM specificities and measures the distance of samples to hyperplane to determine the uncertain values. Third, committee-based approaches such as maximum disagreement (MD)-based criteria (Zhou et al., 2016) which calculates sample uncertainty by considering the incompatible assumption between each committee.

Recently, different diversity approaches based on the closest support vector (Wang et al., 2017), angle (Demir et al., 2011) and clustering (Demir et al., 2011) have been considered to reduce the correlation among the uncertain samples. One of the methods that utilizes both MCLU uncertainty and angle-based diversity (ABD) of data to select the diverse informative samples is MCLU-ABD (Demir et al., 2011).

In this paper, a new multi-criteria AL method which is called spectral-spatial augmented AL (SSAAL) has been proposed. It includes two main steps:

1) The spectral-spatial features are extracted by applying 2-D non-subsampled shearlet transform (NSST) (Lim, 2010) and kernel minimum noise fraction (KMNF). The 2D-NSST is applied to each spectral band of HSI to extract the spatial features. After that, KMNF reduces the spectral dimensions.

2) A new multi-criteria batch mode augmented AL algorithm is applied to the selected spectral-spatial features in order to determine the most informative samples. These samples are selected by a query function which is a combination of a new posterior probability-based uncertainty and an angle-based diversity criteria. Data augmentation (DA) is also considered to increase the informative samples for insufficient labeled classes. DA is iteratively applied to the selected samples during the training process and significantly increases the supervised classification accuracy. The outlier samples are defined by k-means clustering and reduced.

The proposed method is compared to the well-known state-of-the-art techniques. The obtained results demonstrate superior classification performance of the proposed method.

The remainder of this paper is organized as follows: section 2 presents the proposed algorithm. The experimental results are discussed in section 3. Concluding remarks are provided in section 4.

2. PROPOSED METHOD

2.1 Spectral-Spatial Feature Extraction Using NSST and KMNF

The NSST contains two kinds of non-subsample filter banks that are iteratively applied to extract spatial features: pyramid and shearing filters. Pyramid filters divide the image into approximate and detail images which are the same size as the original image for a predefined level of decomposition. Shearing filters decompose the detail images into a number of shearing directional sub-bands. Also, at each iteration of the process, the obtained low-frequency sub-band is again divided into a lower scale high-frequency and low-frequency sub-bands (Soleimanzadeh et al., 2018).

In this paper, 2D-NSST is applied to each band of $X \in R^{I_1 \times I_2 \times I_3}$ as the HSI with the length of I_1 , width of I_2 and I_3 bands. Shearlet transformation of the b th band of the image is defined as:

$$X_{NSST}(:, :, b) = NSST(X(:, :, b)), \quad b = 1, 2, \dots, I_3 \quad (1)$$

After that, in order to reduce the spectral dimension, one of the most popular nonlinear dimensionality reduction techniques "KMNF" (Gao et al., 2017) is used. It consists of two consecutive principle component analysis (PCA) transformations. At first, the prior one estimates the covariance matrix of the noise in the data based on a strong relationship between adjacent pixels to decorrelate and rescale the noise from the data. Then, the latter is a standard PCA transformation which is applied to the matrix and arranges the bands with respect to the signal to noise (Priyadarshini et al., 2019). KMNF is applied to X_{NSST} as follows:

$$X_{KMNF} = KMNF(X_{NSST}) \quad (2)$$

In the proposed method the important shearlet coefficients that contain 99% of the energy are preserved and the rest of them are discarded.

2.2 Augmented AL Algorithm

The obtained spectral-spatial features of the previous step are used as the input of augmented AL algorithm. These features are randomly divided into three categories: initial training set L , unlabeled pool U , and testing set T . During the augmented AL process, at each iteration of algorithm, a batch of m most informative samples among the samples in U is selected by a query function, it is added to the augmented initial training set LA and trained until the maximum number of iterations is achieved. In the following, the algorithm will be explained in more detail.

2.2.1 Uncertainty Criterion A one-against-all SVM is simultaneously used to determine uncertainty criterion and perform the supervised classification (Melgani and Bruzzone, 2004). The class membership probability-based uncertainty (CMPU) criterion is applied to the unlabeled samples in U to define the samples which are confused with the class membership. The uncertainty measure of each unlabeled sample $x \in U$ is related to its classification confidence. For this aim, c binary SVM classifier corresponding to c information classes is iteratively applied to the labeled samples. After training, based on the c obtained decision hyperplanes of SVM, c functional Euclidean distances $f_q(x)$, $q = 1, 2, \dots, c$ for each sample to c hyperplanes are calculated. By utilizing Platt scaling (Lin et al., 2007) and fitting a Sigmoid model to these functional distances, the probability of class membership for each x is calculated. The produced class probabilities in this way can be denoted as:

$$P(y|x) = \frac{1}{1 + \exp(af_q(x) + b)}_{q=1, \dots, c} \quad (3)$$

$$P(y'|x) = \frac{1}{1 + \exp(af_{q'}(x) + b)}_{\substack{q'=1, \dots, c \\ q' \neq q, y' \neq y}}$$

where $P(y|x)$ defines the best class membership probability of sample x and $P(y'|x)$ denotes the second best class membership probability of it. They are the class probabilities between the most confused classes y and y' . a and b are two scalar parameters.

By measuring the ratio of these values, CMPU criterion is obtained. This criterion defines classification confidence of samples and the ones which have the lowest value of it are selected as the most uncertain samples. CMPU is defined as:

$$CC(x) = \frac{P(y|x)}{P(y'|x)} \quad (4)$$

2.2.2 Diversity Criterion It is used to select the samples with low redundancy among n uncertain samples. In this paper, ABD in the kernel space (Demir et al., 2011) is used to calculate the diversity of uncertain sample as follows:

$$ABD(x_i, x_j) = \frac{K(x_i, x_j)}{\sqrt{\prod_{i,j=1, \dots, n} K(x_i, x_i) K(x_j, x_j)}} \quad (5)$$

where $K(x_i, x_j)$ is a nonlinear kernel map function between two samples x_i and x_j .

2.2.3 Query Function In the proposed method, a weighted query function based on integration CMPU and ABD criteria for defining a batch of m ($m < n$) most informative samples is considered.

$$F(x_1, x_2, \dots, x_m) = \arg \min \sum_{i=1}^m \left\{ \lambda [CC(x_i)] + (1 - \lambda) \sum_{j=1}^n \left[\max_{i \neq j} (ABD(x_i, x_j)) \right] \right\} \quad (6)$$

By exploiting objective function (6), the most informative samples (x_1, x_2, \dots, x_m) are extracted at each iteration of SSAAL algorithm. In this function, the parameter λ , ($0 \leq \lambda \leq 1$) controls the relation importance of the two terms and the optimal value of it for each dataset is utilized. The details of selecting optimal λ are discussed in section 3.2.

2.2.4 Data Augmentation (DA) Transformation such as scaling, flipping and rotating are used to increase the insufficient labeled samples during the training process. In the proposed technique, to reduce the running time and prevent overfitting, a new approach of DA is applied to the initial training samples and the most informative samples that are selected at each iteration by query function (6). For this purpose, the non-selective samples of HSI become zero and a new hyper-cube is created. Then, different augmentation methods are applied to the hyper-cube and the obtained non-zero pixels are extracted and added to the training samples. In addition, the same distribution of the original and the augmented samples is an important point that should be considered. Therefore, the two-sample Kolmogorov-Smirnov test (Labadi et al., 2014) is considered for selecting the most effective augmentation methods. It measures the difference between cumulative distribution functions (CDFs) of two existing training and augmented sample vectors and uses the maximum absolute value of it. It can be defined as follows:

$$d = \sup |F_1(x_t) - F_2(x_a)| \quad (7)$$

where x_t and x_a are training and augmented sample vectors with sizes t and a , respectively. $F_1(x_T)$ and $F_2(x_A)$ are their corresponding CDFs.

If d is lower than the critical value $d_{\alpha, t}$ (Knuth, 2014) at a significant level $\alpha = 0.05$, x_t and x_a belong to the same distribution and the augmented samples can be added to the training samples. This critical value can be denoted as:

$$d_{\alpha, t} = \sqrt{-\ln\left(\frac{\alpha}{2}\right) \cdot \frac{1 + \frac{t}{a}}{2t}} \quad (8)$$

Some of the augmented samples are outliers. These samples have a high uncertainty. However, they can decrease the classification accuracy. Therefore, in the proposed method k-means

clustering is used to reduce the outlier effect (Wu and Prasad, 2016). This method allocates samples to k clusters and removes the samples that belong to the smallest cluster by considering equation (9):

$$CR_1 = \frac{C_1}{C_2}$$

$$ACR = \frac{1}{k-2} \sum_{r=2}^{k-1} C_r \quad (9)$$

$$th = \frac{CR_1}{ACR}$$

In this equation, $C = [C_1, C_2, \dots, C_k]$ is an ascending order set that represents the number of cluster members for k cluster and the members of the smallest cluster are outliers if only $th < 0.5$.

The pseudo-code of SSAAL algorithm is summarized in the Algorithm 1.

3. EXPERIMENTS

3.1 Datasets

The proposed technique has been applied to the three real datasets.

3.1.1 Indian Pines Scene (IP) It is acquired by the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) sensor in 1992. This dataset is composed of 220 spectral bands with wavelength varying $0.4-2.5\mu m$ and 145×145 pixels with a spatial resolution $20m \times 20m$. In our experiment 20 bands due to noises and water absorption are eliminated, resulting in 200 bands. This dataset contains 10249 total available samples belonging to 16 classes (see Table 1). The false color composition of this dataset as well as ground reference classification map are shown in Fig. 1.

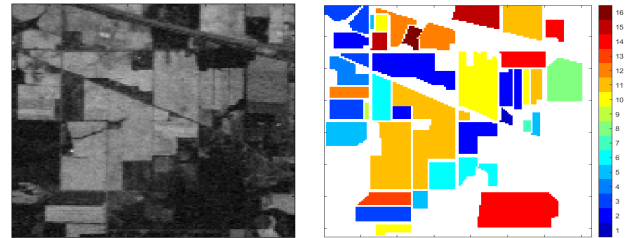


Figure 1. Hyperspectral Indian pines image and its ground-truth map.

3.1.2 Kennedy Space Center (KSC) It is collected over the KSC, Florida, USA, in 1996 (see Fig. 2). This dataset includes a spatial coverage of 512×614 pixels with a spatial resolution of $18m$ and 224 spectral band in the wavelength range from 0.4 to $2.5\mu m$. By removing water absorption and low signal-to-noise bands, the number of spectral bands is reduced to 176. This dataset contains the total number of 5211 samples belonging to 15 classes (see Table 2).

3.1.3 Pavia University Dataset (PU) It is acquired by the Reflective Optics System Imaging Spectrometer (ROSIS) sensor over the urban area of the University of Pavia, Italy, 2003. This dataset consists of 610×340 pixels with a spatial resolution of $1.3m$ (see Fig. 3). The number of spectral bands is 115 with spectral ranging from 0.43 to $0.86\mu m$. 12 noisy bands are removed; the remaining 103 spectral bands are used in the experiment. Table 3 depicts the number of total available pixels in each class in Pavia University dataset.

Algorithm 1: SSAAL Technique

Inputs: HSI dataset $X = \{x_1, x_2, \dots, x_h\}$, Number of shearing directions, Batch size (m), Number of iterations (it), Number of clusters (k), CMPU parameters (a,b)

I) Extract spectral-spatial features:

- a) Apply 2D-NSST to each spectral band of HSI.
- b) Apply KMNF to reduce the spectral dimension.

II) Augmented AL process:

- a) Generate the initial training labeled set L , unlabeled pool U and testing set T .
- b) Augment L and preserve the original and augmented samples in LA .
- c) For $i=1:it$
 - 1: Train LA with SVM.
 - 2: For each $x \in U$ compute its CMPU value.
 - 3: Select n samples from U that have the lowest CMPU.
 - 4: Select a batch of m most informative samples from the $n(= 10m)$ samples using query function (6).
 - 5: Specify labels to the m selected samples.
 - 6: Augment the m obtained samples considering the selected method by two-sample Kolmogorov-Smirnov test.
 - 7: Allocate m selected samples and their augmentations to k clusters and remove the outliers considering th .
 - 8: Include non-outlier samples into LA .
 - 9: Remove the selected samples in step 4 from U .
- d) end

Output: Classification results of HSI $\hat{Y} = \{\hat{y}_1, \hat{y}_2, \dots, \hat{y}_n\}$

Class no.	Class name	No. total samples
1	Alfalfa	46
2	Corn-notill	1428
3	Corn-min	830
4	Corn	237
5	Grass/pasture	483
6	Grass/trees	730
7	Grass/pasture-mowed	28
8	Way-windrowed	478
9	Oats	20
10	Soybeans-notill	972
11	Soybeans-min	2455
12	Soybean-clean	593
13	Wheat	205
14	Woods	1265
15	Buildings-grass-tree-drives	386
16	Stone-steel towers	93
Total		10249

Table 1. Indian pines dataset: class numbers, class names, and number of observations for each class

Class no.	Class name	No. total samples
1	Scrub	761
2	Willow swamp	243
3	Cabbage palm hammock	256
4	Cabbage palm/oak hammock	252
5	Slash pine	161
6	Oak/broadleaf hammock	229
7	Hardwood swamp	105
8	Graminoid marsh	431
9	Spartina marsh	520
10	Cattail marsh	404
11	Salt marsh	419
12	Mud flats	503
13	Water	927
Total		5211

Table 2. KSC dataset: class numbers, class names, and number of observations for each class

3.2 Experimental Setup

Results of the proposed method are compared with two batch mode state-of-the-art AL algorithms: MUCLU-ABD (Demir et al., 2011) and BvSB (Li and Zhang, 2016), and one well-known spectral-spatial AL technique MPM-LBP-BT (Li et al., 2013).

For a fair comparison, the same spectral-spatial feature selection algorithm for MCLU-ABD, BvSB, and the proposed method based on NSST-KMNF is considered. In this experiment, NSST

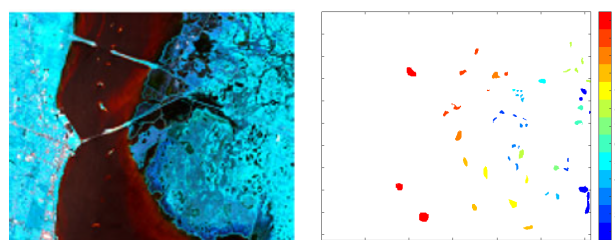


Figure 2. Hyperspectral Indian pines image and its ground-truth map.

Class no.	Class name	No. total samples
1	Asphalt	6631
2	Meadows	18649
3	Gravel	2099
4	Trees	3064
5	Metal sheets	1345
6	Bare soil	5029
7	Bitumen	1330
8	Self-blocking bricks	3682
9	Shadows	947
Total		42776

Table 3. Pavia university dataset: class numbers, class names, and number of observations for each class

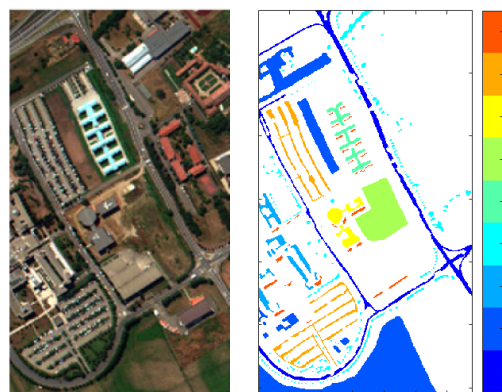


Figure 3. Hyperspectral Pavia University image and its ground-truth map.

with five levels decompositions is used. Also, the important shearlet sub-bands that contain 99% of energy are preserved. For, Indian Pines, KSC, and Pavia University, the number of

shearing directions at four scales and the number of remaining shearlet sub-bands after applying KMNF are presented in Table 4.

Dataset	No. shearing directions	No. remaining shearlet sub-bands
Indian Pines	[2 2 4 4]	72
KSC	[2 2 2 2]	67
Pavia University	[2 2 2 2]	50

Table 4. NSST-KMNF setup of three datasets

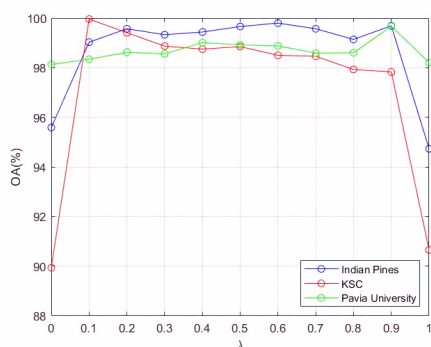


Figure 4. Overall classification accuracy versus the sampling weight parameter (λ)

The same initial training set (L), unlabeled pool (U), testing set (T), number of iterations (it), and batch size (m) are considered for all AL algorithms.

The total samples of each dataset are considered as T . L is constructed by randomly selecting three samples per class and the remaining samples of each dataset are preserved in the U . At each iteration of AL process, m samples from U are selected, they are augmented by DA methods and added to the augmented initial training set LA . Also, the number of clusters k are equal to the number of classes for each dataset plus one. The SSAAL parameter settings of three datasets are demonstrated in Table 5.

The effect of the sampling weight parameter λ in (6) is evaluated in terms of overall classification accuracy (see Fig. 4). As is shown, the optimal value of λ for each dataset is selected among a range of $\lambda = [0, 0.1, 0.2, \dots, 1]$. The optimal values show that although both uncertainty and diversity criteria have important roles on the obtained classification accuracy, the importance of each of them for each dataset is different.

The AL classification algorithms were implemented in MATLAB (R2020a) on a computer with two core processor (2.60 GHz), 40 Gb of memory, and a 64-bit operating system.

LIBSVM is adopted to implement SVM. One against all (OAA) SVM with radial basis function (RBF) kernel has been used (Chang and Lin, 2011). The SVM hyperparameters C and γ were optimized by applying a grid search according to a three-fold cross-validation technique. To obtain better results, these parameters were updated once during the AL iterations. The overall accuracy (OA), average accuracy (AA), Kappa coefficient (Kappa), and classification accuracy of each class are calculated to evaluate the classification performance.

3.3 Experimental Results

Figure 5 depicts the overall classification accuracy for three datasets. In this experiment, the initial accuracy achieved by

SSAAL is higher than the other techniques for Indian Pines and University of Pavia datasets, while KSC obtains the best accuracy after three iterations. Tables 6-8 summarize the classification accuracy, overall accuracy, average accuracy, Kappa coefficient, and computation time. The classification maps of all methods are shown in Figs. 6-8. From Tables 6-8, it can be easily observed that the SSAAL achieves the best performance in terms of AA and OA measures in comparison to MCLU-ABD (Demir et al., 2011), BvSB (Li and Zhang, 2016), and MPM-LBP-BT (Li et al., 2013) methods. In most of the classes, the classification accuracy of SSAAL is higher than the other techniques.

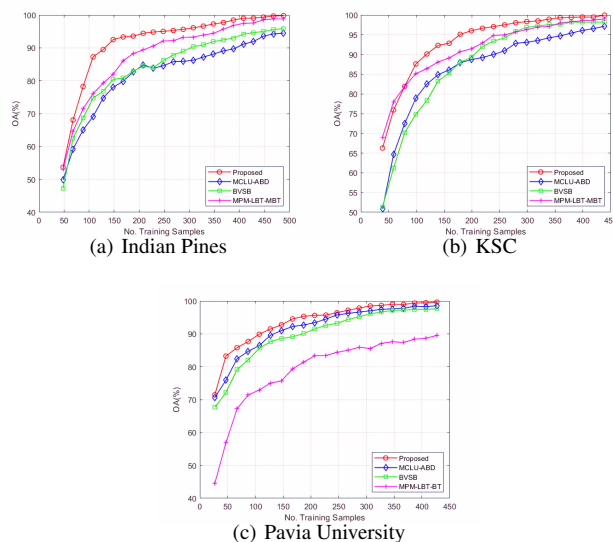


Figure 5. Overall classification accuracy versus the number of training samples.

The computation time of SSAAL method is lower than MPM-LBP-BT and close to MCLU-ABD and BvSB techniques.

The obtained results show that utilizing the combination of NSST and KMNF in the proposed method can extract spectral and spatial features of HSI with low redundancy, effectively. Additionally, the integration of the extracted features, the operative query function in exploiting the most informative and distinct samples, optimal data augmentation during the training process, and outlier elimination can lead to accurate and consistent classification results, significantly.

4. CONCLUSIONS AND FUTURE WORK

In this paper, a new augmented AL algorithm for spectral-spatial HSI classification with limited labeled samples has been presented. The proposed method extracts the spatial features using NSST and reduces the dimensionality of the spectral features using KMNF. After that, a multi-criteria batch mode AL method based on a new query function which integrates CMPU uncertainty and ABD diversity criteria is applied to the extracted spatial-spectral features to select the most informative ones. The determined samples are iteratively trained by SVM. It is note worthy to mention that the selected training samples are augmented using the determined methods by two-sample Kolmogorov-Smirnov test and the existing outliers are removed by k-means clustering. Experiments on three real HSI datasets are performed for validation. Based on the visual and qualitative results, in the proposed method the classification accuracy is significantly enhanced in comparison with three state-of-the-art AL algorithms.

Datasets	No. L samples	No. U samples	No. T samples	No. total training samples	NO. LA samples	No. iterations (it)	DA	Batch size (m)	No. clusters (k)	Sampling weight parameter (λ)
IP	48	10201	10249	488	1948	22	rotating 90°, 180°, flipping vertically	20	17	0.6
KSC	39	5172	5211	439	1728	20	rotating 180°, flipping horizontally, vertically	20	14	0.1
PU	27	42749	42776	427	1708	20	rotating 180°, flipping horizontally, rescaling to 0.9	20	10	0.9

Table 5. Parameter settings for the three HSI datasets

Class	MCLU-ABD	BvSB	MPM-LBP-BT	SSAAL
1	97.83	97.83	95.24	97.83
2	87.61	94.26	95.51	99.72
3	91.33	95.42	99.74	99.40
4	91.98	91.14	100	100
5	97.93	94.82	100	100
6	97.67	98.22	100	100
7	100	100	100	100
8	100	99.58	100	100
9	100	100	100	100
10	94.03	93.93	99.89	100
11	93.12	94.79	98.14	99.55
12	97.81	95.45	100	99.49
13	100	99.02	99.50	100
14	99.68	98.81	99.92	100
15	93.78	96.37	99.72	100
16	100	100	93.10	100
OA(%)	94.50 (0.86)	95.83 (0.33)	98.79 (0.21)	99.80 (0.05)
AA(%)	96.42	96.85	98.80	99.75
Kappa	0.937	0.953	0.988	0.998
Time(min)	9.58	9.06	15.63	10.17

Table 6. Classification accuracies and computation time obtained by different AL techniques on Indian Pines dataset.

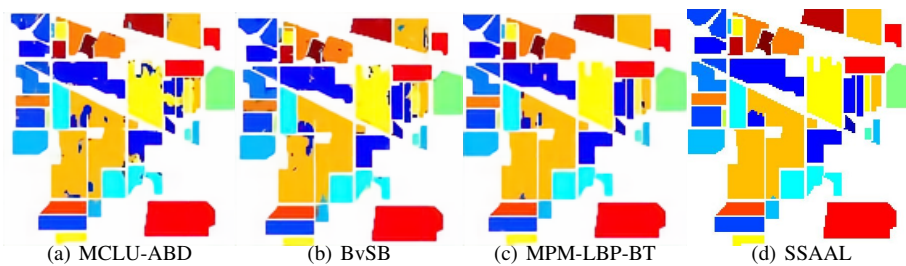


Figure 6. Classification maps obtained by different AL techniques on the Indian Pines dataset.

Class	MCLU-ABD	BvSB	MPM-LBP-BT	SSAAL
1	97.50	90.67	99.45	100
2	100	100	100	100
3	95.31	97.66	100	100
4	82.14	100	86.67	100
5	98.14	100	100	100
6	82.53	94.32	100	99.13
7	72.38	100	100	100
8	100	99.07	100	100
9	100	100	100	100
10	100	100	100	100
11	100	100	95.53	100
12	100	100	100	100
13	100	100	100	100
OA(%)	97.16 (0.34)	98.20 (0.18)	98.99 (0.1)	99.96 (0.03)
AA(%)	94.46	98.59	98.59	99.93
Kappa	0.968	0.980	0.989	0.999
Time(min)	9.62	9.16	196.29	10.66

Table 7. Classification accuracies and computation time obtained by different AL techniques on KSC dataset

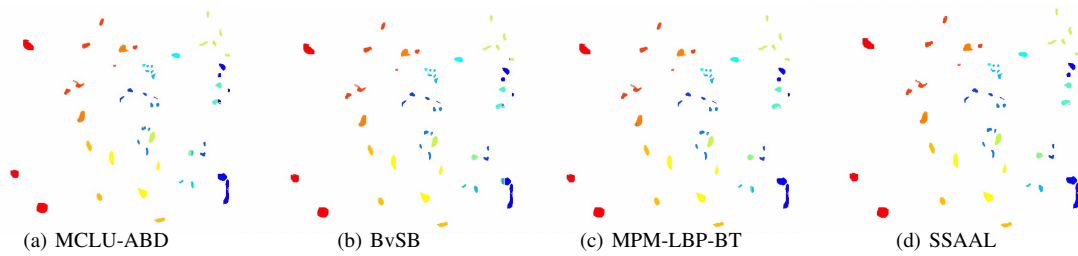


Figure 7. Classification maps obtained by different AL techniques on KSC dataset.

Class	MCLU-ABD	BvSB	MPM-LBP-BT	SSAAL
1	99.16	98.51	96.17	99.73
2	98.36	99.44	96.39	99.75
3	94.66	95.52	51.02	98.81
4	98.24	96.31	92.04	99.05
5	99.93	99.85	94.28	99.93
6	99.80	95.82	92.04	99.98
7	99.62	96.99	14.87	99.77
8	98.07	91.44	81.76	98.29
9	99.58	97.04	97.67	99.68
OA(%)	98.56 (0.37)	97.65 (0.6)	89.52 (3.89)	99.67 (0.02)
AA(%)	98.60	96.77	79.58	99.44
Kappa	0.981	0.969	0.860	0.994
Time(s)	11.26	11.46	128.25	13.20

Table 8. Classification accuracies and computation time obtained by different AL techniques on Pavia University dataset.

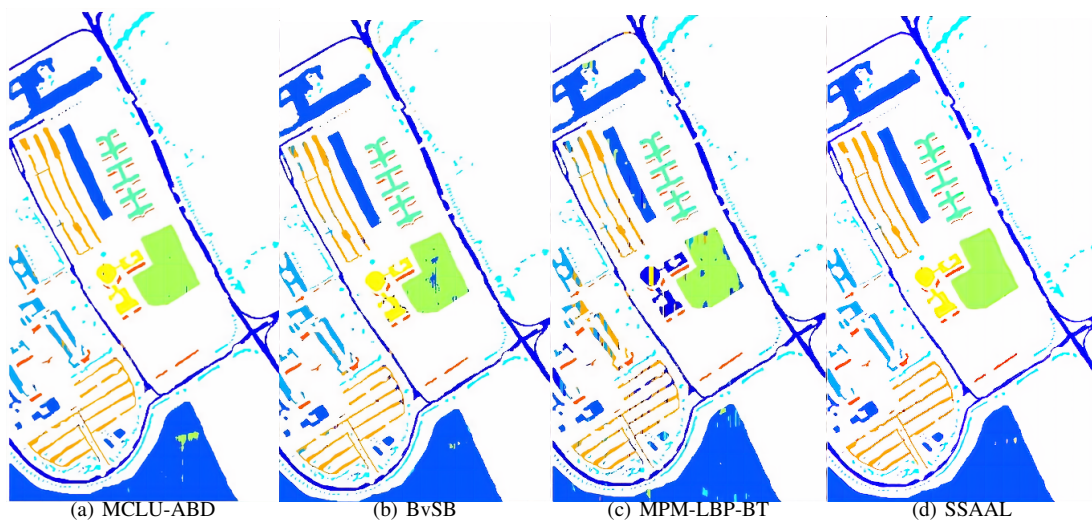


Figure 8. Classification maps obtained by different AL techniques on Pavia University dataset.

REFERENCES

- Ahmad, M., Khan, A., Khan, A. M., et al., 2019. Spatial prior fuzziness pool-based interactive classification of hyperspectral images. *Remote Sensing*, 11(9), 1136.
- Bai, J., Yuan, A., Yu, W., et al., 2020. Decoupled Network With Active Learning Strategy For Hyperspectral Image Classification. 509–512.
- Camps-Valls, G., Tuia, D., Bruzzone, L., et al., 2014. Advances in Hyperspectral Image Classification: Earth Monitoring with Statistical Learning Methods. *IEEE Signal Processing Magazine*, 31(1), 45–54.
- Chang, C.-C., Lin, C.-J., 2011. LIBSVM: A Library for Support Vector Machines. *ACM Trans. Intell. Syst. Technol.*, 2(3).
- Demir, B., Persello, C., Bruzzone, L., 2011. Batch-Mode Active-Learning Methods for the Interactive Classification of Remote Sensing Images. *IEEE Transactions on Geoscience and Remote Sensing*, 49(3), 1014–1031.
- Gao, L., Zhao, B., Jia, X., Liao, W., Zhang, B., 2017. Optimized kernel minimum noise fraction transformation for hyperspectral image classification. *Remote sensing*, 9(6), 548.
- Haut, J. M., Paoletti, M. E., Plaza, J., et al., 2018. Active learning with convolutional neural networks for hyperspectral image classification using a new bayesian approach. *IEEE Transactions on Geoscience and Remote Sensing*, 56(11), 6440–6461.
- He, Z., Liu, H., Wang, Y., et al., 2017. Generative adversarial networks-based semi-supervised learning for hyperspectral image classification. *Remote Sensing*, 9(10), 1042.

- Jun, G., Ghosh, J., 2008. An Efficient Active Learning Algorithm with Knowledge Transfer for Hyperspectral Data Analysis. 1, I-52-I-55.
- Knuth, D. E., 2014. *Art of computer programming, volume 2: Seminumerical algorithms*. Addison-Wesley Professional.
- Labadi, L. A., Masuadi, E., Zarepour, M., 2014. Two-sample Bayesian nonparametric goodness-of-fit test. *arXiv preprint arXiv:1411.3427*.
- Li, C., Zhang, L., 2016. Hyperspectral image classification by combination of active learning and extended multi-attribute profile. 541–544.
- Li, J., 2015. Active learning for hyperspectral image classification with a stacked autoencoders based neural network. 1–4.
- Li, J., Bioucas-Dias, J. M., Plaza, A., 2010. Semisupervised hyperspectral image segmentation using multinomial logistic regression with active learning. *IEEE Transactions on Geoscience and Remote Sensing*, 48(11), 4085–4098.
- Li, J., Bioucas-Dias, J. M., Plaza, A., 2011a. Hyperspectral image segmentation using a new Bayesian approach with active learning. *IEEE Transactions on Geoscience and Remote Sensing*, 49(10), 3947–3960.
- Li, J., Bioucas-Dias, J. M., Plaza, A., 2011b. Spectral-spatial hyperspectral image segmentation using subspace multinomial logistic regression and Markov random fields. *IEEE Transactions on Geoscience and Remote Sensing*, 50(3), 809–823.
- Li, J., Bioucas-Dias, J. M., Plaza, A., 2013. Spectral-Spatial Classification of Hyperspectral Data Using Loopy Belief Propagation and Active Learning. *IEEE Transactions on Geoscience and Remote Sensing*, 51(2), 844–856.
- Lim, W., 2010. The Discrete Shearlet Transform: A New Directional Transform and Compactly Supported Shearlet Frames. *IEEE Transactions on Image Processing*, 19(5), 1166–1180.
- Lin, H.-T., Lin, C.-J., Weng, R. C., 2007. A note on Platt's probabilistic outputs for support vector machines. *Machine learning*, 68(3), 267–276.
- Liu, P., Zhang, H., Eom, K. B., 2016. Active deep learning for classification of hyperspectral images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 10(2), 712–724.
- Melgani, F., Bruzzone, L., 2004. Classification of hyperspectral remote sensing images with support vector machines. *IEEE Transactions on geoscience and remote sensing*, 42(8), 1778–1790.
- Paoletti, M. E., Haut, J. M., Plaza, J., et al., 2020. Training Capsnets Via Active Learning For Hyperspectral Image Classification. 40–43.
- Patra, S., Bruzzone, L., 2011. A Fast Cluster-Assumption Based Active-Learning Technique for Classification of Remote Sensing Images. *IEEE Transactions on Geoscience and Remote Sensing*, 49(5), 1617–1626.
- Persello, C., Bruzzone, L., 2014. Active and Semisupervised Learning for the Classification of Remote Sensing Images. *IEEE Transactions on Geoscience and Remote Sensing*, 52(11), 6937–6956.
- Priyadarshini, K. N., Sivashankari, V., Shekhar, S., Balasubramani, K., 2019. Comparison and evaluation of dimensionality reduction techniques for hyperspectral data analysis. *Multidisciplinary Digital Publishing Institute Proceedings*, 24number 1, 6.
- Rajan, S., Ghosh, J., Crawford, M. M., 2008. An Active Learning Approach to Hyperspectral Data Classification. *IEEE Transactions on Geoscience and Remote Sensing*, 46(4), 1231–1242.
- Soleimanzadeh, M. R., Karami, A., Scheunders, P., 2018. Fusion of Hyperspectral and Lidar Images Using Non-Subsampled Shearlettransform. 8873–8876.
- Sun, S., Zhong, P., Xiao, H., et al., 2015. An MRF model-based active learning framework for the spectral-spatial classification of hyperspectral imagery. *IEEE journal of selected topics in signal processing*, 9(6), 1074–1088.
- Tuia, D., Ratle, F., Pacifici, F., et al., 2009. Active learning methods for remote sensing image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 47(7), 2218–2232.
- Wang, R., Wang, X.-Z., Kwong, S., et al., 2017. Incorporating diversity and informativeness in multiple-instance active learning. *IEEE transactions on fuzzy systems*, 25(6), 1460–1475.
- Wu, H., Prasad, S., 2016. Dirichlet process based active learning and discovery of unknown classes for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 54(8), 4882–4895.
- Yu, H., Sun, C., Yang, W., Yang, X., Zuo, X., 2015. AL-ELM: One uncertainty-based active learning algorithm using extreme learning machine. *Neurocomputing*, 166, 140–150.
- Zhou, X., Prasad, S., Crawford, M. M., 2016. Wavelet-domain multiview active learning for spatial-spectral hyperspectral image classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 9(9), 4047–4059.