



HYPERSPECTRAL IMAGE CLASSIFICATION USING UNCERTAINTY AND DIVERSITY BASED ACTIVE LEARNING

USHA PATEL*, HARDIK DAVE† AND VIBHA PATEL‡

Abstract. There has been extensive research in the field of Hyperspectral Image Classification using deep neural networks. The deep learning based approaches requires huge amount of labelled data samples. But in the case of Hyperspectral Image, there are less number of labelled data samples. Therefore, we can adopt Active Learning combined with deep learning based approaches to be able to extract most informative data samples. By using this technique, we can train the classifier to achieve better classification accuracies with less number of labelled data samples. There is considerable amount of research carried out for selecting diverse data samples from the pool of unlabeled data samples. We present a novel diversity-based Active Learning approach utilizing the information of clustered data distribution. We incorporate diversity criteria with Active Learning selection criteria and combine it with Convolutional Neural Network for feature extraction and classification. This approach helps us in obtaining most informative and diverse data samples. We have compared our proposed approach with three other sampling methods in terms of classification accuracies, Cohen Kappa score, which shows that our approach gives better results with comparison to other sampling methods.

Key words: Active Learning (AL), Convolutional Neural Network (CNN), Hyperspectral Image (HSI) Classification, Deep Learning (DL), Diversity, Uncertainty

AMS subject classifications. 68T07

1. Introduction. Hyperspectral Imaging or Imaging Spectroscopy is a field of collection, analysis and processing of information from across wide electromagnetic spectrum [14]. For this, various instruments called imaging spectrometers are used such as AVIRIS (Airborne Visible Infra-Red Imaging Spectrometer), ROSIS (Reflective Optics System Imaging Spectrometer) etc. [27]. These spectrometers collect large amount of information at different wavelength to collectively analyse the region or scene. Usually these information at different wavelengths are called bands or spectral bands or spectral channels. Hyperspectral images (HSIs) contains hundreds of nearly continuous or narrow spectral bands each of which represents the image at different wavelength, thus it provides rich and detailed information of a particular region or scene. With the rapid development in remote sensing field, hyperspectral imaging has proven its usefulness in various domains such as geosciences, agriculture etc. [14], [31], [27].

Significant amount of research has been carried out in the field of classification of hyperspectral images. The large amount of spectral band provided by Hyperspectral images given better accuracy in classification of these images. Essentially, in HSI classification, every single pixel vector of image is categorized into certain class. HSI classification methods have two major categories, spectral-information-based classification methods and the spectral-spatial-information-based classification methods. Spectral-information-based HSI classification utilizes the spectral information contained by Hyperspectral images to classify each pixel vector in the image. Moreover, instead of utilizing all the spectral bands for classification, techniques like principal component analysis and linear discriminant analysis has been used to reducing the spectral bands and retaining most relevant bands [22], [2]. Spectral-based methods do not utilize spatial correlation information which results in low performance of classifier. Spectral-spatial-information-based methods make use of both the spectral and spatial information by using patch instead of single pixel [2].

Unsupervised techniques or clustering algorithms have also been applied for HSI classification [30]. Supervised classification techniques are mostly preferred in HSI classification because of their capabilities in classifi-

*Nirma University, Ahmedabad, India (ushapatel@nirmauni.ac.in)

†Nirma University, Ahmedabad, India

‡Vishwakarma Government Engineering College, GTU, India

cation with higher performance [13]. Traditional machine learning algorithms such as Support Vector Machines has been employed for HSI classification but did not give significant performance [25]. Various approaches based upon deep learning algorithms have also been applied for HSI classification. Deep belief networks in [20] and [8], Stacked Auto-Encoders (SAE) in [26] and Convolutional Neural Networks (CNN) in [16], [6] and [15] have been used and were successful. CNN are very effective approach for Hyperspectral image classification. CNN is designed in such a way that it can extract features in the images in one go and with less computation. In [6], 1D, 2D and 3D CNNs have been used to classify HSIs utilizing its spectral as well as spatial information. Spectral-spatial-based approach of using 3D CNN gives much better results compared to other methods.

However, these deep learning based approaches require huge amount of labelled data which is very difficult, infeasible and expensive to collect in practice. To address this issue, some methods have been proposed using data augmentation techniques which is used to generate additional data [21]. More efficient method to tackle this issue is Active Learning (AL) which selects most informative samples for labelling by the oracle and thus reducing number of samples required for the classifier to gain significant performance [29]. Active Learning selects highly informative or most useful samples for the current model to learn from using a query function. By including only those samples selected by active learning query function we significantly reduce the amount of labelled data needed for the model to reach certain performance mark.

Active Learning strategies assume that not all the data samples are equal but few samples have characteristics that separates the classes. The idea is to select only those samples for training and still keep or even enhance the performance of the model. AL has been extensively studied for HSI classification in [28], [32], [33].

Diversity in the data samples while active selection has been less studied. Most studies in AL based HSI classification considers only most informative samples for inclusion in training set. This can increase redundancy in data samples to be trained degrading performance of the classifier. Thus, the query function employed for selection of the most useful or informative or uncertain samples should consider two major criteria: 1) informativeness and 2) diversity of data [3], [36]. The diversity criterion selects data samples from a set of most uncertain unlabelled data samples which are as diverse as possible, therefore reducing redundancy among the data samples. Combining the diversity criterion with the query function of the AL technique, the efficiency of the model will be increased potentially.

In this paper, a method to incorporate the AL technique with diversity criterion and deep learning is proposed to utilize effectiveness of all the fields. We propose new clustering-based diversity criterion along with the AL query function which will be incorporated with 3D CNNs to utilize the strong discriminative capabilities of deep learning. By using the diversity criterion with AL query function, the model becomes more robust and generalizes well due to the training of the model on different and diverse data samples. Although, there have been several studies which combines deep learning with AL but our proposed approach has different characteristics which is discussed in related work. Moreover, we employ the data augmentation technique to produce new data samples which helps in addressing the difficulty of insufficient labelled data samples. Furthermore, we retrain the CNN model only on the actively selected data instead of training on whole data, which reduces the computational cost of training the CNN model on whole data again. We have compared four different AL strategies to measure effectiveness of our proposed method. We have performed experiments on three well-known datasets to validate our proposed method.

2. Related Work.

2.1. Deep Learning based HSI classification. Due to the rapid growth in deep learning algorithms research and their applicability, researchers have been attracted to use deep learning algorithms for solving HSI classification problems. Several studies have been carried out in this domain. In [20], [8], Deep Belief Networks were used for feature extraction and classification. In the paper [20], spectral-spatial-based classification was also considered in experimentation. SAE (Stacked Auto-Encoders) which can be formed by multiple sparse autoencoders were used for HSI classification in [26] and features in the data are learned in an unsupervised way. There has been significant work in HSI classification using CNN compared to DBN and SAE. CNNs are more popular because of its ability to learn features in the images naturally [18]. In [16], five layered basic CNN model was proposed for HSI classification, considering only spectral information. Several works has been carried out to consider spectral-spatial features also. In [6], 3D patch-based CNN was used considering spectral-spatial features and in [15] B-CNN (Bayesian CNN) were employed with comparison of 1D, 2D and

3D CNN architectures. In [21], data augmentation techniques were employed for improve the performance of the CNN model.

2.2. Active Learning based HSI classification. Active Learning has been extensively used in HSI classification in recent years. Active Learning strategies can be divided into three major categories 1) Committee-based strategies 2) Large margin-based strategies and 3) Posterior probability based strategies [33]. In [9], the entropy was used as a query function or heuristic combine with Query-by-committee approach. Large margin-based heuristics are mostly used with support vector machines (SVM) due to their direct applicability with margin of separating hyperplane [7] [33]. In [5], [4], naïve and modified version of Margin Sampling (MS) was used as an active learning query function with SVM which is based on selection of samples within the margin of the hyperplane. In [11], the query function MS is extended for multiclass problem with SVM and the heuristic is called multiclass level uncertainty (MCLU). MCLU considers two most probable cases instead of considering only one like MS. Posterior probability based approaches make use of posterior probabilities estimates of individual class membership. Several studies employ the KL-Max heuristic and Breaking ties heuristic in the posterior probability based approaches [17], [33].

Active learning incorporated with deep learning has been actively studied in recent years. In [23], Deep-Transfer learning approach has been employed which is incorporated with AL. Representativeness of the data samples are examined by considering them simultaneously in [23]. In [1], discriminative CNN (D-CNN) were used with AL to consider diversity in the data samples. Similar to [23], study in [12] also considers deep learning and transfer learning together with AL and stacked sparse autoencoders (SSAE) were used. In [6], HSI classification has been carried out using CNNs with AL criteria BvSB. This study also considers Markov Random Fields (MRF) for class label smoothness. The authors of [15] have proposed Bayesian CNN (B-CNN) based framework for HSI classification. Six different active learning strategies were employed and compared in [15].

2.3. Diversity based HSI classification. While selecting the most uncertain or most informative data samples, it is more probable that the selected data samples are redundant among each other. The data samples should be diverse from the current training data also to avoid redundancy. These two approaches namely, 1) diversity among each other and 2) diversity from current training data or current data distribution [33]. In [33], the authors have surveyed the need of diversity criterion and avoiding redundancy. In [29], MS is extended to MS-cSV (closest Support Vectors) to consider the distribution of the data samples in the multidimensional space of features to select only those data samples residing in the different area or region, therefore preserving the diversity. In [3], the selection criteria incorporates diversity among the samples by considering angles between the induced hyperplanes. The authors of [11] have proposed two diversity criteria namely angle based diversity (ABD) which is similar to the work in [3] and clustering based approach based on kernel k-means combined with AL query function. Clustering based method to preserve diversity in has been employed by clustering the uncertain samples and then selecting only one most uncertain or informative sample per cluster [11]. In [34], cluster based approach for diversity in the data samples was employed in different way. The cluster with highest number of uncertain data samples is further divided into two sub-clusters, until certain predefined number of clusters q , the number of samples to be selected for retraining, has been generated. This approach is quite expensive in terms of computation.

In [10], density is also considered along with diversity. Clustering is done on uncertain samples and most representative sample having lowest average distances from all the points in the cluster is selected. In [35], the diversity is found by using a maximum mean discrepancy, a measure for calculating the difference between two different datasets in terms of data distribution. In [19], the authors have proposed the nearest neighbour approach for preserving the diversity.

3. Proposed Approach. In our proposed approach, we have considered a novel clustering based method for preserving the diversity of the data samples from the current training data which boosts the model performance and makes model more robust and generalized.

We also incorporate the diversity among the data samples in the initial training phase for enhancement of the discriminative abilities of the model. In our proposed method DB-MCLU (Diversity-based MCLU), we have combined Diversity based criteria and BvSB (Best versus Second Best). We have employed the BvSB (Best versus Second Best) strategy for actively selecting samples [6]. Also, we have incorporated diversity

criteria along with BvSB to preserve diversity among the samples which are actively selected using BvSB. The BvSB strategy for active learning is based on MCLU (Multiclass Level Uncertainty) [33]. The MCLU strategy considers two classes with the highest class membership probabilities and the difference of distances of both to the margin in case of SVM classifier [33]. Similar approach is used in BvSB strategy where posterior probabilities are computed and then difference between two highest class membership probabilities are used as a measure of uncertainty. The data samples with least difference of highest two probabilities are considered as most informative and uncertain data samples. The small value of BvSB measure denotes that the model is not certain with the data sample's class membership and thus the data sample is uncertain. For a data sample x_i , the BvSB measure can be calculated using equation 3.1:

$$BvSB(x_i) = P_B(x_i) - P_{SB}(x_i) \quad (3.1)$$

The diversity criteria in the proposed approach is implemented by considering clustered data distribution of the current training data and then by selecting only those actively selected data samples which are diverse from the current training data. Clustering of the current training data is done to measure diversity of actively selected samples from current data distribution. A novel clustering based method is used for preserving the diversity in actively selected data samples. First, the current training data is clustered using well known K-Means clustering algorithm in C clusters where C is the number of classes in the dataset [24]. The clusters in the K-Means clustering algorithm can be formed using the equation 3.2 for samples (x_1, x_2, \dots, x_n) . Where each sample is a d -dimensional vector, then k clusters $C = \{C_1, C_2, \dots, C_k\}$ of these samples can be formed.

$$\operatorname{argmin}_c \sum_{i=1}^k \sum_{x \in C_i} \|x - \mu_i\|^2 \quad (3.2)$$

The Euclidean distance for two points a and $b \in R^d$ can be calculated using following formula:

$$d(a, b) = \sqrt{\sum_{i=1}^d (b_i - a_i)^2} \quad (3.3)$$

Then, a cluster is assigned to each of the actively selected data samples based on the minimum Euclidean distance from a cluster's centroid i.e. the distances from each cluster centroid to a particular sample x_i are calculated and a cluster's centroid from which the distance to x_i is minimum, is assigned to that particular data sample x_i . Then for those samples who were assigned the same cluster, average of distances from those samples to the assigned cluster's centroid is calculated and the data samples whose distance to the assigned cluster's centroid is greater than the average distance calculated before, are chosen as diverse samples. This step in the proposed method preserves the diversity in the actively selected samples from current training data distribution.

We have employed data augmentation methods for images such as Flip Up-Down, Flip Left-Right and Rotate the image to increase the amount of labelled data samples. The hyperspectral image contains hundreds of spectral bands which may lead to over-fitting problem. So, to extract most representative dimensions from the whole dataset, we make use of dimensionality reduction method such as Principal Component Analysis (PCA). Instead of retraining the CNN model on whole training data in every iteration, we retrain the CNN model only on the data samples selected by uncertainty and diversity criteria. The data samples at each iteration are selected for retraining the CNN model based on combination of uncertainty and diversity criteria. The complete structure of the proposed approach is shown in Fig. 3.1. In the Fig. 3.1, the word fine-tuning refers only to retraining of CNN model with only the actively selected data and not the whole dataset. This fine-tuning has nothing to do with structural/architectural changes in CNN model.

The distribution of whole dataset is shown in Table 3.1.

In Algorithm 1, the initial training of the CNN model is carried out on D_{train} . The remaining dataset is D_{remain} . The 60% of (D_{remain}) is called D_{pool} from which we actively select samples for retraining of CNN. The 40% of the D_{remain} is called D_{test} which will be utilized for evaluation of the trained CNN model. D_{active} is the data which is selected from D_{pool} for retraining on the basis of active learning and diversity criteria.

We have designed the Algorithm 2 to be executed until desired criteria is met. The desired criteria can be Time limit, Model Accuracy criteria or it can be Number of actively selected samples for training. We have

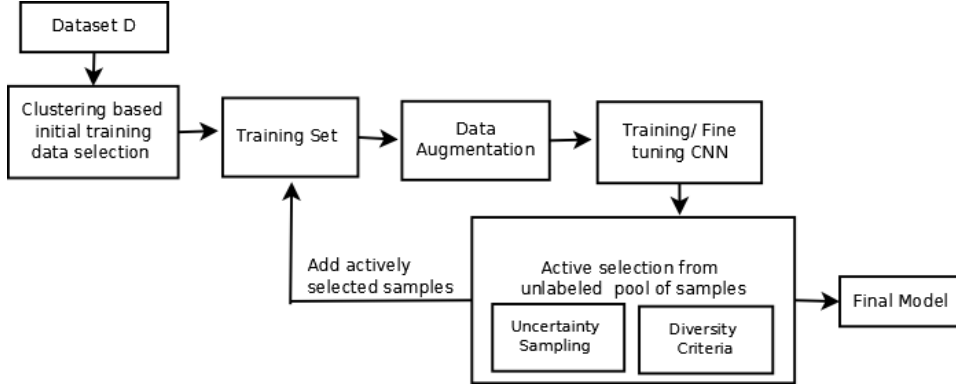


FIG. 3.1. Proposed diversity based HSI classification framework

TABLE 3.1
Samples for Each Dataset

Dataset	Train	Pool	Test	Total
Indian pines	400	5909	3940	10249
Salinas	400	32237	21492	54129
Pavia University	225	25530	17021	42776

selected accuracy criteria for implementation which means we execute the algorithm until we reach desired model accuracy. The data samples are selected for retraining from the D_{pool} based on DB-MCLU approach. The DB-MCLU approach combines diversity criteria along with MCLU (i.e. based on BvSB). According to Algorithm 2, "h" most uncertain data samples are selected from D_{pool} based on BvSB Active Learning criteria. Further, from these data samples, some data samples D_{active} are selected based on diversity criteria. Thus we ensure to combine both uncertainty and diversity in selection of data samples for retraining. This approach preserves diversity among the uncertain samples.

4. Implementation Details and Results. The proposed approach is implemented on three publicly available Hyperspectral datasets (Table 3.1) and discussed implementation perspective in this section.

4.1. Dataset Description. The first HSI dataset used is well known Indian Pines dataset which was collected by AVIRIS (Airborne Visible/Infrared Imaging Spectrometer) sensor in 1992. This dataset has 145 x 145 pixels with 220 spectral bands with range of the wavelength between 0.4 and 2.5 μm . Each pixel in this image dataset is assigned one of the 16 ground-truth classes. The second dataset used is Pavia University which was collected by ROSIS (Reflective Optics System Imaging Spectrometer) sensor. The dataset consists 610 x 340 pixels with nine land-cover classes and 103 spectral bands with the wavelength range between 0.43 and 0.86 μm . The third dataset used in experiments was Salinas Valley scene which was collected by AVIRIS sensor. The dataset consists of 512 x 217 pixels and has 204 bands after removing 20 water absorption bands. The dataset has 16 ground-truth classes.

4.2. Hyperparameter Settings and Description. In the experiments, we have two major hyperparameters which are empirically tuned. The first hyperparameter "s" is number of samples to be selected from each cluster for initial training data. This hyperparameter is carefully tuned to be 25 during experiments. The number of initial training samples greatly affect on the final classification results because subsequent iterations of active selection of uncertain and informative data samples depend greatly on the initially trained CNN model. The second hyperparameter "h" is the number of most uncertain and informative data samples selected by active learning strategy. This hyperparameter is tuned empirically to 200. The diversity based criteria is employed on these actively selected data samples and less than half of the data samples from these 200 data samples are selected, which was observed during experiments. The other hyperparameters for CNN training are set empirically such as batch size in [16, 32] and number of epochs in the range of [10, 20]. We

Algorithm 1: Initialize D_{train} , D_{pool} , D_{remain} , D_{test} **Input:** Initial Dataset D **Output:** Initial D_{train} , D_{pool} , D_{remain} , D_{test} Where N number of classes, t split ratio, s number of samples selected

- Run K-Means clustering algorithm on initial Dataset and get clusters $\{C_i\}_{i=1}^N$ and cluster centroids $\{\Delta_i\}_{i=1}^N$ from equation (2)
- Select s samples from each cluster based on minimum Euclidean distance given in equation (3)
- $D_{train} = \{x_i, y_i\}_{i=1}^l$ where $l = s * N$
- $D_{remain} = D / D_{train}$, split D_{remain} into D_{pool} and D_{test} with ratio t

Algorithm 2: Training using Active Learning (DB-MCLU)**Input:** Initial D_{train} , D_{pool} , D_{remain} , D_{test} **Output:** Trained Model with D_{train} SamplesTrain/Retrain CNN model on current training dataset D_{train} **repeat** $D_{uncertain} = \{x_i\}_{i=1}^h$ Where, h is most uncertain samples from D_{pool} using BvSB AL strategyUpdate $\{C_i\}_{i=1}^N$ and $\{\Delta_i\}_{i=1}^N$ based on current D_{train} Assign a cluster C_j to sample x_i based on minimum distance from Δ_j to the sample x_i For each,
 $x_i \in D_{uncertain}$ **for** $x_i \in C_j$ **do**Find average distances a_j

$$a_j = \frac{\sum_{i=1}^r d(\Delta_j, x_i)}{r}$$

where, r is number of samples in each clusterAdd sample x_i to D_{active} where, $d(\Delta_j, x_i) > a_j$ $D_{pool} = D_{pool} / D_{active}$ $D_{train} = D_{train} \cup D_{active}$ Retrain the CNN model with D_{active} and Get prediction on D_{test} **until** desired criterion;

used dimensionality reduction technique PCA to obtain most useful and representative dimensions of dataset and to decrease the complexity. We consider 30 principal components for Indian Pines dataset and 15 principal components for rest of the two datasets. The value of principal components is chosen based on the variance of the dataset. We preserved 99.6% variance in the datasets after applying PCA. To preserve the variance of 99.6% in the dataset practically, the values 30 and 15 principal components were chosen for Indian Pines and Rest of the two datasets respectively. The learning rate is set to 0.001.

For all the experiments carried out, we used Google Colab platform for all the computation work.

4.3. Comparison with other Active Learning Methods: In this section, the comparison of our proposed method and three other Active Learning methods based on uncertainty sampling strategy is presented. These three other Active Learning methods are namely, 1) Random Sampling or Baseline, 2) BvSB and 3) Maximum Entropy.

1. Random Sampling or Baseline: The Random Sampling or Baseline method for choosing data samples employs the uniform random distribution for selection of the data samples. This method chooses the data samples following uniform random distribution from pool of data samples by generating random number over the interval $[0,1]$.

2. BvSB (Best versus Second Best): Best versus Second Best method for selection of the uncertain data samples is based on the MCLU strategy [6], [33]. It considers difference between highest and second highest class membership probabilities for selection criteria according to equation – (1). Less the difference of highest and second highest class membership probabilities, more uncertain and informative the data sample is.
3. Maximum Entropy: This method selects data samples having highest uncertainty in classification [6], [29]. In other words, this method selects the data samples that maximizes the entropy. The maximum entropy for a data sample x_i can be calculated by using the following equation:

$$E[y_i|x_i, D_{train}] = - \sum p(y_i = c|x_i, D_{train}) \log p(y_i = c|x_i, D_{train}) \quad (4.1)$$

4. DB-MCLU: Our proposed method is different from these three methods described above. These three methods have only single criteria for selection of the data samples for consequent training. In contrast, our method combines two criteria for selection of data samples, one based on uncertainty and the second based on diversity. For uncertainty, we use BvSB approach as it is most suitable for our model and gives efficient results than Random Sampling and Maximum Entropy measure. We incorporate diversity criteria also with the uncertainty criteria to build the model more robust and generalized. The diversity criteria in our method is based on clustering of current training data, associating the actively selected samples with one of the clusters based on the Euclidean distances and then selecting distant data samples from current training data distribution. By integrating diversity criteria in selection of data samples for subsequent retraining iteration for CNN model, we preserve dissimilarity and variety in data samples.

4.4. Results and Discussion. Series of experiments are conducted on three datasets using four different AL strategies. As we consider diversity in the selection of initial training set also, we use K-means clustering algorithm with number of clusters taken as number of classes for particular dataset. Then we select 25 samples nearest to the centroid of each cluster. Thus we get 400 diverse samples in case of Indian Pines and Salinas dataset and 225 samples in case of Pavia University dataset as our initial training dataset. We split remaining data into D_{pool} and D_{test} using split ratio 0.4 i.e. from the remaining data 60% data samples were used as pool and 40% data samples were used as testing set. As per this split ratio, the partitioned data samples for training, testing and pool are shown in Table 3.1.

For each iteration of selecting the data samples actively, we choose data samples in the range of [50, 200] from D_{pool} based on the uncertainty criteria. This range has been observed empirically. The second criteria for diversity reduces these selected samples by considering only diverse samples from current data distribution. We repeat the active learning iterations until we reach desired accuracy on testing dataset. The average experimental results are shown in Table 4.1, 4.2 and 4.3. For achieving the accuracy given in Table 4.1 for Indian Pines dataset, only 15% of the total data samples have been used. Similarly, for Pavia University and Salinas dataset only 5% of total data samples are used for achieving given accuracies in Table 4.2 and 4.3. From these observations, we can say this method significantly reduces the number of labelled samples required for achieving higher accuracy. The experimental results of all the methods are shown in the Table 4.1, 4.2 and 4.3 for three datasets respectively. The results of using other active learning methods are compared with the proposed method in the Table 4.1, 4.2 and 4.3 in terms of OA (Overall Accuracy), AA (Average Accuracy), CA (Class Accuracy) and K (Cohen Kappa Score). For all the methods, the experiments are repeated five times and the average of all the results are shown. In the experiments conducted on Indian Pines dataset shown in Table 4.2, the methods Entropy and Random Sampling did not able to correctly classify or distinguish the samples for class “Oats”. The number of data samples belong to the class ”Oats” are only 20 in the whole dataset. Even then the methods BvSB and DB-MCLU gave better classification accuracy than other methods with very less data samples of a particular class.

From Table 4.1, 4.2 and 4.3, it can be observed that the proposed method DB-MCLU achieves best classification results with measures of OA (Overall Accuracy) and AA (Average Accuracy) and also, for most of the classes, the proposed method gives best class accuracies compared to other methods. Classification result obtained by our proposed approach along with ground truth image for all three dataset is shown in Figure 4.1.

TABLE 4.1
Classification Accuracies on Indian Pines Dataset

Class/Method	Random	EP	BvSB	DB-MCLU
Alfalfa	54.44	61.11	93.33	88.89
Corn-notill	90.41	89.05	93.53	93.15
Corn-mintill	91.35	95.77	88.04	93.74
Corn	95.27	89.03	96.99	95.48
Grass-pasture	93.72	87.23	96.38	96.17
Grass-trees	97.80	95.53	97.16	96.17
Grass-pasture-mowed	0	83.63	76.36	87.27
Hay-windrowed	100	90.05	99.45	99.78
Oats	0	0	45	65
Soybean-notill	88.31	94.85	87.69	94.79
Soybean-mintill	96.73	96.94	94.01	94.82
Soybean-clean	84.73	84.64	84.91	92.41
Wheat	98.63	92.27	89.86	96.99
Woods	98.26	97.02	99.29	97.39
Buildings-Grass-Trees-Drives	96.16	89.35	87.55	97.35
Stone-Steel-Towers	100	83.17	83.70	91.11
AA	80.36	82.12	88.33	92.53
OA	93.52	92.79	93.09	95.08
K	92.60	91.78	92.11	94.39

TABLE 4.2
Classification Accuracies on Salinas Dataset

Class/Method	Random	EP	BvSB	DB-MCLU
Brocoli_green_weeds_1	99.80	99.95	99.92	100
Brocoli_green_weeds_2	99.93	99.33	99.04	95.54
Fallow	93.37	97.52	98.43	99.85
Fallow_rough_plow	83.73	97.99	93.01	100
Fallow_smooth	73.39	96.66	80.83	95.66
Stubble	99.34	99.67	97.91	98.95
Celery	97.48	97.89	96.85	99.79
Grapes-untrained	97.89	94.16	92.64	97.75
Soil_vinyard_develop	99.97	99.93	97.15	99.91
Corn_senesced_green_weeds	98.70	97.43	96.13	99.25
Lettuce_romaine_4wk	96.30	91.88	95.22	99.49
Lettuce_romaine_5wk	96.43	97.27	85.94	99.87
Lettuce_romaine_6wk	86.79	94.26	77.39	91.67
Lettuce_romaine_7wk	85.03	92.98	67.31	89.01
Vinyard_untrained	93.25	93.32	93.79	93.65
Vinyard_vertical_trellis	99.78	96.18	99.91	99.97
AA	93.82	96.66	91.97	97.52
OA	95.64	96.58	93.83	97.63
K	95.14	96.19	93.13	97.36

The method DB-MCLU achieves best accuracies as compared to other shown methods while using very less number of labelled data samples, 15% of total data samples for Indian Pines and only 5% of total data samples for Pavia University and Salinas Dataset. Thus our proposed method helps in reducing the cost for labelling the data samples while still preserving the diversity of data samples and higher performance of the model.

5. Conclusion. In this paper, we have presented a novel diversity-based Active Learning approach for classification of Hyperspectral Images using Convolutional Neural Network. We have incorporated the diversity criteria with Active Learning sampling strategy. Thus we can consider the diversity in the data samples in

TABLE 4.3
Classification Accuracies on Pavia University Dataset

Class/Method	Random	EP	BvSB	DB-MCLU
Shadows	68.83	59.58	83.27	69.95
Self-Blocking Bricks	90.91	94.18	86.95	95.94
Bitumen	94.83	96.46	91.62	89.33
Bare Soil	98.37	99.70	94.95	98.25
Painted metal sheets	98.90	100	59.85	93.86
Trees	92.69	91.30	89.37	95.27
Gravel	82.30	88.12	61.82	94.02
Meadows	99.29	98.82	99.06	99.61
Asphalt	95.25	96.56	96.39	98.14
AA	91.27	91.64	84.81	92.55
OA	95.70	96.20	92.81	97.13
K	94.40	94.97	90.41	96.19

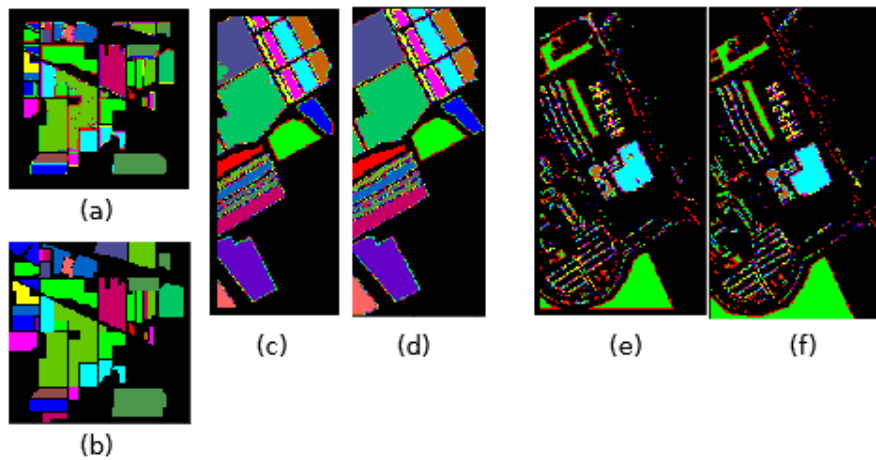


FIG. 4.1. (a), (c), (e) Classification result of Indian Pine, Salinas and Pavia University. (b), (d), (f) Ground truth image of Indian Pine, Salinas and Pavia University.

each subsequent round of active selection of most informative data samples from the pool of unlabelled data samples to be labelled by oracle. We also make use of data augmentation technique to leverage the training process as it provides more information by augmenting the data samples. The retraining of the Convolutional Neural Network model only on the actively selected data samples instead of training on whole dataset significantly reduces the training cost. The proposed diversity-based Active Learning approach incorporating with Convolutional Neural Network for feature extraction and classification tasks effectively utilizes the strengths of all the techniques together and gives better classification performance compared to other sampling methods used for Hyperspectral Image Classification.

The future work includes incorporating different novel diversity based approaches apart from clustering and exploring more Active Learning selection strategies or design a novel one. This work can be extended to semi-supervised and unsupervised approach utilizing the abilities of both the domain.

REFERENCES

- [1] G. ARULAMPALAM, *A generalised feedforward neural network architecture and its applications to classification and regression*, (2004).
- [2] T. V. BANDOS, L. BRUZZONE, AND G. CAMPS-VALLS, *Classification of hyperspectral images with regularized linear discriminant analysis*, *IEEE Transactions on Geoscience and Remote Sensing*, 47 (2009), pp. 862–873.

- [3] K. BRINKER, *Incorporating diversity in active learning with support vector machines*, in Proceedings of the 20th international conference on machine learning (ICML-03), 2003, pp. 59–66.
- [4] L. BRUZZONE AND C. PERSELLO, *Active learning for classification of remote sensing images*, in 2009 IEEE International Geoscience and Remote Sensing Symposium, vol. 3, IEEE, 2009, pp. III–693.
- [5] C. CAMPBELL, N. CRISTIANINI, A. SMOLA, ET AL., *Query learning with large margin classifiers*, in ICML, vol. 20, 2000, p. 0.
- [6] X. CAO, J. YAO, Z. XU, AND D. MENG, *Hyperspectral image classification with convolutional neural network and active learning*, IEEE Transactions on Geoscience and Remote Sensing, (2020).
- [7] X. CEAMANOS, B. WASKE, J. A. BENEDIKTSSON, J. CHANUSSOT, M. FAUVEL, AND J. R. SVEINSSON, *A classifier ensemble based on fusion of support vector machines for classifying hyperspectral data*, International Journal of Image and Data Fusion, 1 (2010), pp. 293–307.
- [8] Y. CHEN, X. ZHAO, AND X. JIA, *Spectral–spatial classification of hyperspectral data based on deep belief network*, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 8 (2015), pp. 2381–2392.
- [9] L. COPA, D. TUIA, M. VOLPI, AND M. KANEVSKI, *Unbiased query-by-bagging active learning for vhr image classification*, in Image and Signal Processing for Remote Sensing XVI, vol. 7830, International Society for Optics and Photonics, 2010, p. 78300K.
- [10] B. DEMIR AND L. BRUZZONE, *A novel active learning method in relevance feedback for content-based remote sensing image retrieval*, IEEE Transactions on Geoscience and Remote Sensing, 53 (2014), pp. 2323–2334.
- [11] B. DEMIR, C. PERSELLO, AND L. BRUZZONE, *Batch-mode active-learning methods for the interactive classification of remote sensing images*, IEEE Transactions on Geoscience and Remote Sensing, 49 (2010), pp. 1014–1031.
- [12] C. DENG, Y. XUE, X. LIU, C. LI, AND D. TAO, *Active transfer learning network: A unified deep joint spectral–spatial feature learning model for hyperspectral image classification*, IEEE Transactions on Geoscience and Remote Sensing, 57 (2018), pp. 1741–1754.
- [13] P. GHAMISI, J. PLAZA, Y. CHEN, J. LI, AND A. J. PLAZA, *Advanced spectral classifiers for hyperspectral images: A review*, IEEE Geoscience and Remote Sensing Magazine, 5 (2017), pp. 8–32.
- [14] A. F. GOETZ, G. VANE, J. E. SOLOMON, AND B. N. ROCK, *Imaging spectrometry for earth remote sensing*, science, 228 (1985), pp. 1147–1153.
- [15] J. M. HAUT, M. E. PAOLETTI, J. PLAZA, J. LI, AND A. PLAZA, *Active learning with convolutional neural networks for hyperspectral image classification using a new bayesian approach*, IEEE Transactions on Geoscience and Remote Sensing, 56 (2018), pp. 6440–6461.
- [16] W. HU, Y. HUANG, L. WEI, F. ZHANG, AND H. LI, *Deep convolutional neural networks for hyperspectral image classification*, Journal of Sensors, 2015 (2015).
- [17] G. JUN AND J. GHOSH, *An efficient active learning algorithm with knowledge transfer for hyperspectral data analysis*, in IGARSS 2008-2008 IEEE International Geoscience and Remote Sensing Symposium, vol. 1, IEEE, 2008, pp. I–52.
- [18] K. KUPPALA, S. BANDA, AND T. R. BARIGE, *An overview of deep learning methods for image registration with focus on feature-based approaches*, International Journal of Image and Data Fusion, (2020), pp. 1–23.
- [19] Y. LENG, X. XU, C. SUN, C. CHENG, H. WAN, J. FANG, AND D. LI, *A svm active learning method based on confidence, knn and diversity*, in 2015 IEEE International Conference on Multimedia and Expo (ICME), IEEE, 2015, pp. 1–6.
- [20] T. LI, J. ZHANG, AND Y. ZHANG, *Classification of hyperspectral image based on deep belief networks*, in 2014 IEEE international conference on image processing (ICIP), IEEE, 2014, pp. 5132–5136.
- [21] W. LI, C. CHEN, M. ZHANG, H. LI, AND Q. DU, *Data augmentation for hyperspectral image classification with deep cnn*, IEEE Geoscience and Remote Sensing Letters, 16 (2018), pp. 593–597.
- [22] G. LICCIARDI, P. R. MARPU, J. CHANUSSOT, AND J. A. BENEDIKTSSON, *Linear versus nonlinear pca for the classification of hyperspectral data based on the extended morphological profiles*, IEEE Geoscience and Remote Sensing Letters, 9 (2011), pp. 447–451.
- [23] J. LIN, L. ZHAO, S. LI, R. WARD, AND Z. J. WANG, *Active-learning-incorporated deep transfer learning for hyperspectral image classification*, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 11 (2018), pp. 4048–4062.
- [24] S. LLOYD, *Least squares quantization in pcm*, IEEE transactions on information theory, 28 (1982), pp. 129–137.
- [25] T. MOUGHAL, *Hyperspectral image classification using support vector machine*, in Journal of Physics: Conference Series, vol. 439, IOP Publishing, 2013, p. 012042.
- [26] A. O. B. ÖZDEMİR, B. E. GEDIK, AND C. Y. Y. ÇETIN, *Hyperspectral classification using stacked autoencoders with deep learning*, in 2014 6th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), IEEE, 2014, pp. 1–4.
- [27] A. PLAZA, J. A. BENEDIKTSSON, J. W. BOARDMAN, J. BRAZILE, L. BRUZZONE, G. CAMPS-VALLS, J. CHANUSSOT, M. FAUVEL, P. GAMBA, A. GUALTIERI, ET AL., *Recent advances in techniques for hyperspectral image processing*, Remote sensing of environment, 113 (2009), pp. S110–S122.
- [28] S. RAJAN, J. GHOSH, AND M. M. CRAWFORD, *An active learning approach to hyperspectral data classification*, IEEE Transactions on Geoscience and Remote Sensing, 46 (2008), pp. 1231–1242.
- [29] B. SETTLES, *Active learning literature survey*, tech. report, University of Wisconsin-Madison Department of Computer Sciences, 2009.
- [30] Y. TARABALKA, J. A. BENEDIKTSSON, AND J. CHANUSSOT, *Spectral–spatial classification of hyperspectral imagery based on partitionial clustering techniques*, IEEE Transactions on Geoscience and Remote Sensing, 47 (2009), pp. 2973–2987.
- [31] M. TEKE, H. S. DEVECİ, O. HALILOĞLU, S. Z. GÜRBÜZ, AND U. SAKARYA, *A short survey of hyperspectral remote sensing applications in agriculture*, in 2013 6th International Conference on Recent Advances in Space Technologies (RAST), IEEE, 2013, pp. 171–176.

- [32] D. TUIA, F. RATLE, F. PACIFICI, M. F. KANEVSKI, AND W. J. EMERY, *Active learning methods for remote sensing image classification*, IEEE Transactions on Geoscience and Remote Sensing, 47 (2009), pp. 2218–2232.
- [33] D. TUIA, M. VOLPI, L. COPA, M. KANEVSKI, AND J. MUNOZ-MARI, *A survey of active learning algorithms for supervised remote sensing image classification*, IEEE Journal of Selected Topics in Signal Processing, 5 (2011), pp. 606–617.
- [34] M. VOLPI, D. TUIA, AND M. KANEVSKI, *Memory-based cluster sampling for remote sensing image classification*, IEEE Transactions on Geoscience and Remote Sensing, 50 (2012), pp. 3096–3106.
- [35] Z. WANG, X. FANG, X. TANG, AND C. WU, *Multi-class active learning by integrating uncertainty and diversity*, IEEE Access, 6 (2018), pp. 22794–22803.
- [36] Z. XU, K. YU, V. TRESP, X. XU, AND J. WANG, *Representative sampling for text classification using support vector machines*, in European conference on information retrieval, Springer, 2003, pp. 393–407.

Edited by: Dana Petcu

Received: Apr 9, 2021

Accepted: Sep 30, 2021