




Article

Spatial Prior Fuzziness Pool-Based Interactive Classification of Hyperspectral Images

Muhammad Ahmad ^{1,2,3,*} , Asad Khan ^{4,†}, Adil Mehmood Khan ^{3,†}, Manuel Mazzara ^{5,†} , Salvatore Distefano ^{1,†}, Ahmed Sohaib ^{2,†} and Omar Nibouche ^{6,†} 

¹ Dipartimento di Matematica e Informatica—MIFT, University of Messina, Messina 98121, Italy; sdistefano@unime.it

² Department of Computer Engineering, Khwaja Fareed University of Engineering and Information Technology, Rahim Yar Khan 64200, Pakistan; ahmed.sohaib@kfueit.edu.pk

³ Institute of Data Science and Artificial Intelligence, Innopolis University, Innopolis 420500, Russia; a.khan@innopolis.ru

⁴ School of Computer Science, South China Normal University, Guangzhou 510000, China; asad@mail.ustc.edu.cn

⁵ Institute of Software Development and Engineering, Innopolis University, Innopolis 420500, Russia; m.mazzara@innopolis.ru

⁶ Faculty of Computing, Engineering and the Built Environment, Ulster University, Newtownabbey, Co Antrim BT37 0QB, UK; o.nibouche@ulster.ac.uk

* Correspondence: mahmad00@gmail.com; Tel.: +92-321-6617922

† These authors contributed equally to this work.

Received: 26 March 2019; Accepted: 5 May 2019; Published: 13 May 2019



Abstract: Acquisition of labeled data for supervised Hyperspectral Image (HSI) classification is expensive in terms of both time and costs. Moreover, manual selection and labeling are often subjective and tend to induce redundancy into the classifier. Active learning (AL) can be a suitable approach for HSI classification as it integrates data acquisition to the classifier design by ranking the unlabeled data to provide advice for the next query that has the highest training utility. However, multiclass AL techniques tend to include redundant samples into the classifier to some extent. This paper addresses such a problem by introducing an AL pipeline which preserves the most representative and spatially heterogeneous samples. The adopted strategy for sample selection utilizes fuzziness to assess the mapping between actual output and the approximated a-posteriori probabilities, computed by a marginal probability distribution based on discriminative random fields. The samples selected in each iteration are then provided to the spectral angle mapper-based objective function to reduce the inter-class redundancy. Experiments on five HSI benchmark datasets confirmed that the proposed Fuzziness and Spectral Angle Mapper (FSAM)-AL pipeline presents competitive results compared to the state-of-the-art sample selection techniques, leading to lower computational requirements.

Keywords: hyperspectral imaging; active learning; fuzziness; spectral angle mapper; soft threshold

1. Introduction

Hyperspectral Imaging (HSI) is a technological tool which takes into account a wide spectrum of light instead of just primary colors such as red, green, and blue to characterize a pixel [1,2]. The light striking a pixel, indeed, is divided into many different spectral bands to provide more information on what is imaged. HSI has been adopted in a wide range of real-world applications including biomedical imaging, geosciences, and surveillance to mention a few [3]. One of the main challenges in the HSI domain is the management of the data, which typically yields hundreds of contiguous and narrow spectral bands with very high spatial resolution throughout the electromagnetic spectrum [2,4].

Therefore, HSI classification is complex and can be dominated by a multitude of urban classes and nested regions, than the traditional monochrome or RGB images [5].

Supervised classification methods are widely adopted in the analysis of HSI datasets. These methods include, for example, multinomial logistic regression [6], random forests [7], ensemble learning [8], deep learning [9], support vector machine (SVM) [10], and k-nearest neighbors (KNN) [11]. However, supervised classifiers often under-perform due to the Hughes phenomenon [12], also known as the issue of dimensionality, which occurs whenever the number of available labeled training samples is considerably lower than the number of spectral bands required by the classifier [11]. Figure 1 (Loss of accuracy in terms of ground maps) and Table 1 (Loss of accuracy in terms of overall and kappa (κ) for different number of labeled training samples i.e., 1% and 10% respectively) illustrates the loss in predictive performance of such classification methods for a particular ground image (Pavia University) when using two different sample size.

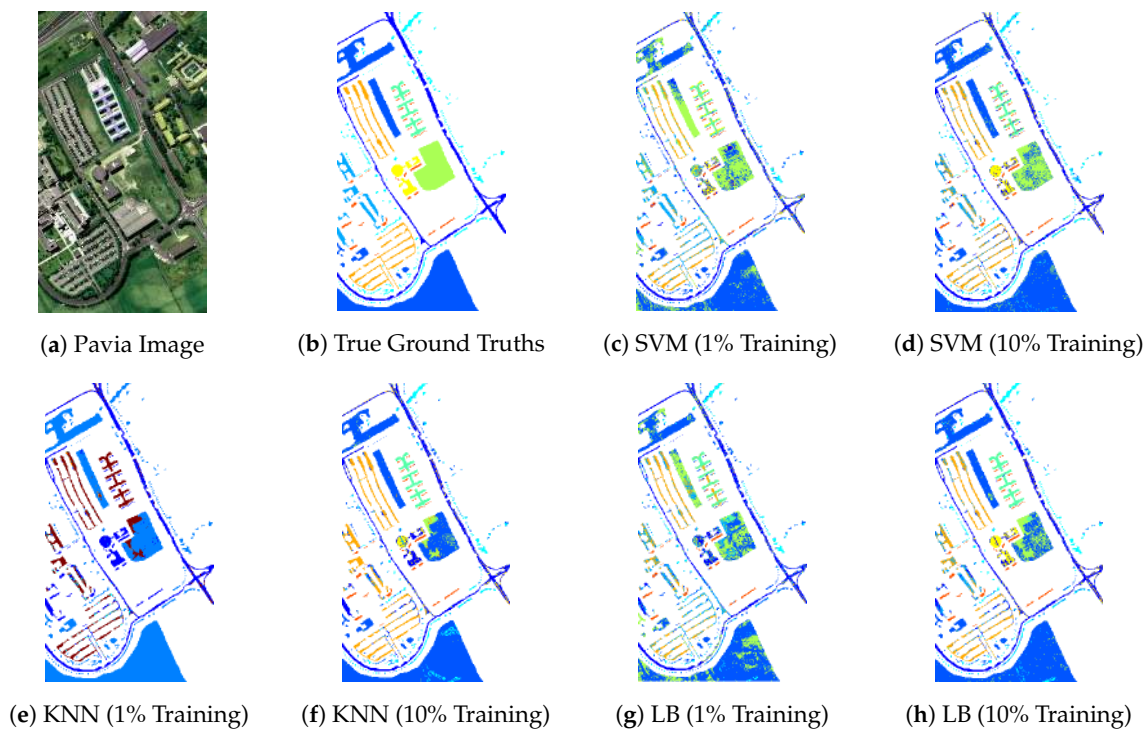


Figure 1. (a) Pavia University image; (b) True ground truths differentiate nine different classes; (c) SVM trained with 1% randomly selected training samples; (d) SVM trained with 10% randomly selected training samples; (e) KNN trained with 1% randomly selected training samples; (f) KNN trained with 10% randomly selected training samples; (g) Logistic boost (LB) trained with 1% randomly selected training samples; (h) LB trained with 10% randomly selected training samples.

Table 1. Classification accuracies in-terms of overall and kappa (κ) obtained by three different classifiers with two different number of randomly selected labeled training samples, i.e., 1% and 10% respectively. All these classifiers are trained and tested using 5-fold cross validation, and from results one can conclude that SVM produce better results when trained with 10% randomly selected training samples. However, the obtained results are not good enough to identify the ground materials accurately, therefore, further investigations are required.

Classifiers	1% Training Samples		10% Training Samples	
	kappa (κ)	Overall	kappa (κ)	Overall
SVM	0.5853%	0.6818%	0.7804%	0.8365%
KNN	0.4897%	0.6488%	0.6791%	0.7691%
LB	0.5216%	0.6488%	0.7531%	0.8365%

The limited availability of labelled training data in the HSI-domain is one of the motivations for the utilization of semi-supervised learning [13]. Examples of such methods include kernel techniques [14] such as SVM, Tri-training [15] algorithms which generates three classifiers from the original labeled samples, then these classifiers are refined using unlabeled samples in the tri-training process, and Graph-based learning [16,17]. A major limitation of such approaches, however, is the low predictive performance when utilizing a small number of training samples within high dimensionality, as commonly observed in HSI classification [18,19] as shown in Figure 1 and Table 1.

Active learning (AL) is a class of semi-supervised learning method which tackles the limitations as mentioned earlier [20,21]. The main component of an AL method is the iterative utilization of the training model to acquire new training samples to be entered to the training set for the next iteration [22]. AL methods can be pool-based or stream-based depending on how they enter new data to the training set, and employ measures like uncertainty, representativeness, inconsistency, variance, and error to rank and select new samples [23]. Despite the gained success, there are still particular characteristics which can cause AL to present inflated false discovery rate and low statistical power [11]. These characteristics include: (i) sample selection bias; (ii) high correlation among the bands; and (iii) non-stationary behavior of unlabeled samples.

This study introduces a customized AL pipeline for HSI to reduce sample selection bias whilst maintaining the data stability in the spatial domain. The presented pipeline distinguishes from standard AL methods in three relevant aspects. First, instead of simply using the uncertainty of samples to select new samples, it utilizes the fuzziness measure associated with the confidence of the training model in classifying those samples correctly. Second, it couples samples' fuzziness with their diversity to select new training samples which simultaneously minimize the error among the training samples while maximizing the spectral angle between the selected sample and the existing training samples. In our current work, instead of measuring angle-based distances among all new samples and all existing training samples, a reference sample is selected from within the training set against which the diversity of the new samples is measured. This achieves the same goal while reducing the computational overhead as the size of training set is always much smaller than the validation set which is the source of new samples. Thirdly, the proposed Fuzziness and Spectral Angle Mapper (FSAM) method keeps the pool of new samples balanced, giving equal representation to all classes, which is achieved via softening the thresholds at run time. Experimental results on five benchmark datasets demonstrate that the customized spatial AL pipeline leads to an increased predictive power regarding kappa (κ) coefficient and overall accuracy, precision, recall, and F1-Score.

The remainder of the paper is structured as follows. Section 2 reviews related work on learning methods applied to HSI. Section 3 presents the theoretical aspects of the proposed FSAM-AL pipeline followed by a theoretical explanation of the implemented objective function. Section 4 discusses the experimental dataset and experimental settings. Section 5 presents the experimental results and intensive discussion on obtained results. Section 6 compare the experimental results with the state of the art sample selection methods used in AL frameworks in the recent most years. Finally, Section 7 summarizes the contributions and future research directions.

2. Related Work

The HSI technology has been employed in several real-world applications including object detection and recognition, change detection (The process of identifying the changes occurred over the time on the earth surface), human-made material identification, semantic annotation, unmixing, and classification [2,16]. However, some challenges can arise from typical characteristics of HSI data, notably the limited availability of labeled data which can lead to inflated false discovery rate and low statistical power [22]. This aspect results in a relatively poor predictive performance of supervised [24] and semi-supervised [25] learning methods when addressing HSI classification, as shown in Figure 1.

AL poses as an alternative for coping with the limited amount of labeled data by iteratively selecting informative samples for the training set [11]. Alternatives of sample selection method utilized

in AL and corresponding references to the literature are shown in Table 2. Table 2 classifies the references in the literature according to the information utilized by the sample selection methods, being either spectral (consider only the wavelength of the pixel) or spectral–spatial (pixel location in addition to the wavelength). The latter class is particularly relevant in the HSI-domain as the acquisition of training samples depends on a large degree on the spatial distribution of the queried samples. However, only a few studies have integrated spatial constraints into AL methods [26,27]. In the Table 2 we have provided a unified summary of existing sample selection methods and the information they use along with references to their respective papers.

Table 2. Different sample selection methods used in Active Learning frameworks for hyperspectral image classification in the recent years.

Sample Selection Methods	References	
	Spectral	Spectral–Spatial
Random selection	[11,20,28]	[29]
Mutual information	[30,31]	[32]
Breaking ties	[33]	[34,35]
Modified breaking ties	[34]	[34,35]
Uncertain sampling	[36,37]	[38–40]
Fisher information ratio	[41]	[42]
Fuzziness information	[11]	—
Query by committee	[43]	—

Tuia et al. [28] presented a detailed survey on AL methods addressing HSI analysis and contrasted non-probabilistic methods, which assume that all query classes are known before to the initialization, to probabilistic approaches that allow the discovery of new classes. The latter class was also pointed as more suitable for cases when the prior assumption is no longer fulfilled [44]. In addition to probabilistic and non-probabilistic AL methods, large margin heuristics have been utilized as the base learner to combine the benefits of HSI analysis and AL [28,45]. A particular approach for selecting samples that have achieved remarkable results for several applications is query by committee (QBC) [46]. Contrarily from previous methods, QBC selects samples based on the maximum disagreement of an ensemble of classifiers. Overall, these sampling methods suffer from high computational complexity due to the iterative training of the classifier for each sample [28].

Pool-based AL, also known as batch-mode AL, addresses the high computational complexity observed in the aforementioned methods by concomitantly considering the uncertainty (spectral information) and diversity (spatial information) of the selected samples [36]. A seminal work was presented by Munoz-Mari et al. [47], which highlighted the benefits of integrating spatio-contextual information to AL even when the distribution of queried samples in the spatial space is ignored. This method was later expanded to include the position of selected samples in the feature space [32]. One of the outcomes from such a transformation is the point-wise dispersed distribution in the spatial domain, which incurs the risk of revisiting the same geographical location several times, especially in the HSI-domain [32].

A considerable amount of research has been conducted on AL in the recent years, often analyzing only spectral properties, whilst ignoring spatial information that plays a vital role in HSI classification as shown in [32,48]. Spatial and spectral HSI classification can achieve higher performance than its pixel-wise counterpart as it utilizes not only information of spectral signature but also from spatial domain [48]. Thus, the combination of spatial and spectral information for AL represents a novel and promising contribution yet to be explored in the HSI-domain. This study proposes a customized pool-based AL pipeline which exploits both spatial and spectral information in the context of HSI classification.

3. Methodology

We address the small sample problem when classifying high dimensional HSI data by defining an AL scheme selecting a pool of diverse samples by taking into account two main criteria. The first is the fuzziness of samples, which is associated with the confidence of the trained model in properly classifying the unseen samples. The second is the diversity of the samples, thus reducing the redundancy among the selected samples. The combination of two criteria results in the selection of a pool of potentially most informative and diverse samples in each iteration.

Although there have been lots of different sampling methods (few mentioned in Table 2), uncertainty remains one of the most popular method that can be used to select the informative samples [11,49]. Usually, the most uncertain samples have similar posterior probabilities for two most possible classes [49]. Thus, probabilistic model could be directly used to evaluate the uncertainty of unlabeled sample [49].

However, assessing the uncertainty of a sample is not as straight-forward when one is using non-probabilistic (NP) classifiers because their output does not exist in the form of posteriori probabilities [23,49]. The output of such classifiers can be manipulated to obtain an approximation of posteriori probability functions for the classes being trained [23].

Suppose $X = [x_1, x_2, x_3, \dots, x_L]^T \in \mathcal{R}^{L \times (M \times N)}$ is an HSI cube which is composed of L spectral bands and $(M \times N)$ samples per band belonging to C classes where $x_i = [x_{1,i}, x_{2,i}, x_{3,i}, \dots, x_{L,i}]^T$ is the i^{th} sample in the cube. Let us assume $(x_i, y_i) \in \mathcal{R}^{(M \times N)} \times \mathcal{R}^C$, where y_i is the class label of the i^{th} sample. Let us further assume that n finite (limited) number of labeled training samples are selected from X to create the training set $D_T = \{(x_i, y_i)\}_{i=1}^n$. The rest of the samples form the validation set $D_V = \{(x_i, y_i)\}_{i=1}^m$. Please note that $n \ll m$, and $(D_T \cap D_V) = \emptyset$.

An NP classifier trained on D_T when tested on D_V would produce an output matrix μ of $m \times C$ dimensions containing NP outputs of the classifier. Let μ_{ij} be the j^{th} output (membership for the j^{th} class) for i^{th} sample. There are several methods proposed in the literature to transform such NP outputs into the posteriori probabilities [23,49]. Such methods are computationally complex in two folds. First these methods need to compute the Bayesian decision for each samples x_i choosing the category y_j having the largest discriminant function $f_j(x_i) = p(\mu_j|x_i)$. Secondly, these methods assume that the training outputs are restricted as $\{0, 1\}$. However, these methods also consider to manipulate each Bayes rule using Jacobin's derivation over the limit theorem on infinite number samples to approximate the posterior probabilities in a least squares sense, i.e., $f(x_i, \mu_j) = p(\mu_j, x_i)$ [49].

In order to overcome the above mentioned difficulties, in this work, we used marginal probability distribution [32] which is obtained from the D_V information in the HSI data, serves as an engine in which our AL pipeline can exploit both the spatial and spectral information in the data. The posteriori class probabilities are modeled with the discriminative random field [32,50] in which the association potential is linked with discriminative, generative, ensemble and signal hidden layer feed forward neural network based classifiers. Thus, the posteriori probabilities are computed as similar to the work [32]. From these posteriori probabilities we obtained the membership matrix which should satisfy the following properties [11]:

$$\sum_{j=1}^C \mu_{ij} = 1 \quad \text{and} \quad 0 < \sum_{i=1}^N \mu_{ij} < 1 \quad (1)$$

In Equation (1), $\mu_{ij} \in [0, 1]$ and $\mu_{ij} = \mu_j(x_i)$ is a function that represents the membership of i^{th} sample x_i to the j^{th} class [11]. For the true class, the posteriori probability would be approximated as close to 1, whereas, if the output is small (wrong class), the probability would be approximated as close to 0. However, AL methods do not require accurate probabilities, but only need a ranking of the samples according to their posteriori probabilities which would help to estimate the fuzziness [49] and the output of the sample.

The fuzziness ($E(\mu)$) upon ($M \times N$) samples for C classes from the membership matrix (μ_{ij}) can then be defined as expressed in Equation (2) which must satisfy the properties defined in [51,52].

$$E(\mu) = \frac{-1}{N \times C} \sum_{i=1}^N \sum_{j=1}^C \left[\mu_{ij} \log(\mu_{ij}) + (1 - \mu_{ij}) \log(1 - \mu_{ij}) \right] \quad (2)$$

Then, we first associate $E(\mu)$, predicted class labels, and actual class labels with D_V and then sort the D_V in descending order based on the fuzziness values. We then heuristically select the \hat{m} number of misclassified samples which have higher fuzziness, where $\hat{m} \ll m$. The proposed strategy keeps the pool of \hat{m} new samples balanced, giving equal representation to all classes, which is achieved via softening the thresholds at run time.

Next, the spectral angular mapper (SAM) (More information about spectral angle mapper (SAM) function can be found in the following papers [53–55]) function is used to discriminate the samples within the same class to minimize the redundancy among the pool of \hat{m} selected samples. SAM is an automated method for directly comparing sample spectra to a known spectra. It treats both spectra as vectors and calculates the spectral angle between them. It is insensitive to illumination since it uses only the vector direction and not the vector length [56]. The output of SAM is an image showing the best match of each pixel at each spatial location [57]. This approach is typically used as a first cut for determining the mineralogy and works well in area of homogeneous regions.

In this work, SAM takes the arc-cosine-based dot product between the test spectrum which have higher fuzziness $D_V^H = \{(x_{ij}^l, y_j)\}_{i=1}^{\hat{m}}$ to a reference (training samples) spectrum $D_T = \{(x_{pj}^l, y_j)\}_{p=1}^n$, where $j \in \{1, 2, 3, \dots, C\}$ and $l \in \{1, 2, 3, \dots, L\}$ where L is the total number of bands in HSI dataset, with the following objective functions:

$$\angle(\alpha_j) = \min_{p \in n} \left(\cos^{-1} \frac{\sum_{l=1}^L x_{pj}^l \cdot x_{ij}^l}{\sqrt{\sum_{l=1}^L (x_{pj}^l)^2} \sqrt{\sum_{l=1}^L (x_{ij}^l)^2}} \right) \quad (3)$$

Equation (3) aims to compute the spectral difference among all the training samples for C classes, respectively. We then select one reference spectrum from each class which minimizes the angular distance among others within same class, i.e., the sample which is more similar to others in the given class. This process will return the number of reference spectrum's up to the number of classes in HSI. We then pick one reference spectrum from $\angle(\alpha_j)$ to compare with all the selected test spectrum for the same class and account the angular distance among them in $\angle(\beta_{ij})$ as shown in Equation (4).

$$\angle(\beta_{ij}) = \sum_{j=1}^C \sum_{i=1}^{\hat{m}} \left(\cos^{-1} \frac{\sum_{l=1}^L (\angle(\alpha_j)) \cdot x_{ij}^l}{\sqrt{\sum_{l=1}^L (\angle(\alpha_j))^2} \sqrt{\sum_{l=1}^L (x_{ij}^l)^2}} \right) \quad (4)$$

$$Ind_{(D_V^H)} = \operatorname{argmax}_{i \in (D_V^H) \setminus \mathcal{X}} \left(\phi(\angle(\beta_{ij})) \right) \quad (5)$$

where $Ind_{(D_V^H)}$ denotes the induces of samples which have higher fuzziness, $D_V^H \setminus \mathcal{X}$ represents the index of samples of D_V^H that are not contained in \mathcal{X} , ϕ provides the trade-off between diversity, and \mathcal{X} denotes the index of the unlabeled sample that will be included in the pool. Please note that here we used a soft threshing scheme to balance the number of classes in both training and selected samples. The proposed pipeline systematically select the $(h / (\text{number of classes}))$ higher fuzziness samples from D_V^H for each class, if one or more classes missed in the pool of selected samples. This process is repeated until the cardinality of \mathcal{X} is equal to h , i.e., $|\mathcal{X}| = h$, where h is the size of pool. This technique guarantees that the selected samples in \mathcal{X} are diverse regarding to their angles to all the others in $(\angle(\beta_{ij}))$. Since the initial size of \mathcal{X} is zero, thus, the first sample included in \mathcal{X} is always the higher fuzziness sample from $E(\mu)$.

There are several advantages of using fuzziness information carried out through SAM as query function: (i) easy to implement; (ii) robust in mapping the spectral similarity for reference to higher fuzziness test spectrum only; (iii) powerful because it represents the influence of shading effects to accentuate the selected test reflectance characteristics [55]. On the other hand, the main drawback of SAM is spectral mixture problems, i.e., SAM assumes that the reference spectrum chosen to classify the HSI represents the pure spectrum which is not the case in our problem. Such problems occur when the HSI is in low or medium spatial resolution. Furthermore, as we know, the surface of the earth is heterogeneous and complex in many ways, thus, containing many mixed samples. The spectral confusion in samples can lead to overestimation or underestimation errors for a spectral signatures. This is not the case of the proposed solution, since we iteratively select the reference spectrum from each class using Equation (3) as a pure spectrum and comparing this with selected test spectrum respectively using Equation (4) with the help of whitening parameter to minimize the redundancy among the selected samples. The complete work-flow of our proposed pipeline is described in Algorithm 1 and Figure 2.

Algorithm 1: Pseudo-code of our Proposed FSAM Algorithm.

```

Data:  $D_T, D_V$  training and test set, respectively.
1 Initialization:  $\mathcal{X}$ ;  $\hat{m}$  = number of samples to select;  $\phi$ ;  $h$ ;
2 while  $|D_T| \leq Threshold$  do
3   Train and Test ELM, SVM, kNN and EL;
4    $\mu_{ij} \leftarrow$  Compute the membership matrix;
5    $E(\mu) \leftarrow$  Compute the fuzziness;
6    $D_V^H \leftarrow$  Associate the fuzziness, actual and predicted class and spatial information with  $D_V$  and sort in descending order;
7   Select  $\hat{m}$  misclassified samples from  $D_V^H$ ;
8    $\angle(\alpha_j) \leftarrow$  Compute the spectral angle between training samples;
9   while  $|\mathcal{X}| \leq h$  do
10     $\angle(\beta_{ij}) \leftarrow$  Compute the spectral angle between reference spectrum and  $\hat{m}$  selected samples;
11     $Ind_{D_V^H} \leftarrow$  Pick the index of most diverse samples.;
12     $\mathcal{X} \leftarrow$  index of selected samples;
13  end
14  Pick  $\mathcal{X}$  samples from  $D_V$ , add them to  $D_T$ , and remove from  $D_V$ ;
15 end

```

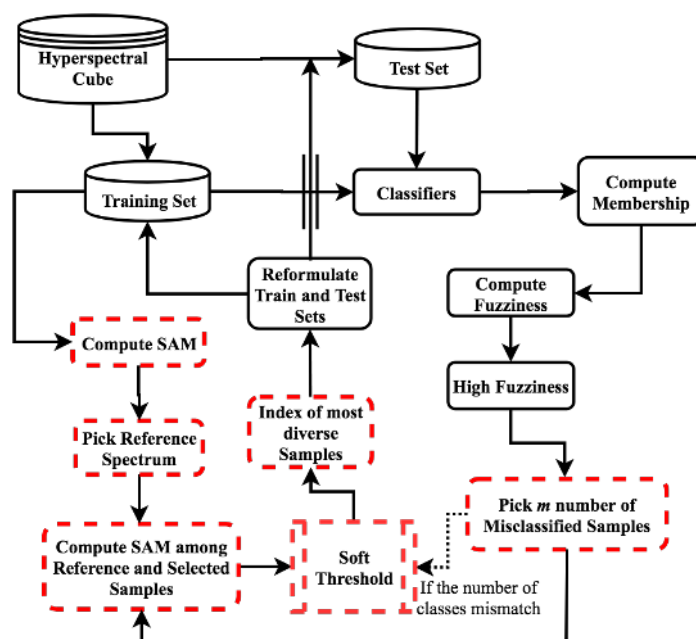


Figure 2. Proposed FSAM-AL Pipeline in which red labeled boxes represent where we contribute.

4. Experimental Datasets and Settings

The performance of our proposed FSAM-AL pipeline is validated on five benchmark HSI datasets acquired by two different sensors, e.g., Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) and Reflective Optics System Imaging Spectrometer (ROSIS). These datasets include, Salinas-A, Salinas full scene, Kennedy Space Center (KSC), Pavia University (PU), and Pavia Center (PC) (Further information about these datasets can be found in [58]).

We evaluated the FSAM pipeline against four different classifiers: extreme learning machine (ELM) [59], support vector machine (SVM), k-nearest neighbor (kNN), and ensemble learning (EL). These classifiers are chosen because they have been extensively studied in the literature for HSI classification and rigorously utilized for comparison purposes. Furthermore, our goal is to show that the proposed method can work well with a diverse set of classifiers. The performance of aforementioned classifiers is measured using two well know metrics: overall accuracy and kappa (κ) coefficient [11]. Furthermore, F1-score, precision, and recall rates are also compared. To further validate the real time applicability of FSAM, we compared it against four benchmark sample selection methods, namely: random sampling (RS), mutual information (MI), breaking ties (BT), and modified breaking ties (MBT).

1. **Random Sampling (RS)** [11,28] method relies on the random selection of the samples without considering any specific conditions.
2. **Mutual Information (MI)** [32] of two samples is a measure of the mutual dependence between the two samples.
3. **Breaking Ties (BT)** [33] relies on the smallest difference of the posterior probabilities for each sample. In a multiclass settings, BT can be applied by calculating the difference between two highest probabilities. As a result, BT finds the samples minimizing the distance between the first two most probable classes. The BT method generally focuses on the boundaries comprising many samples, possibly disregarding boundaries with fewer samples.
4. **Modified Breaking Ties (MBT)** [34,35] includes more diversity in the sampling process as compared to BT. The samples are selected by maximizing the probability of the largest class for each individual class. MBT takes into account all the class boundaries by conducting the sampling in cyclic fashion, making sure that the MBT does not get trapped in any class whereas BT could be trapped in a single boundary.

In all experiments, the initial training size is set as 100 samples from an entire HSI data. In each iteration the size of training set increases with $h = 1\%$ actively selected samples by FSAM pipeline. The best part of FSAM is that there is no hyper-parameters need to be tuned except classification methods. In ELM, the hidden neurons are systematically selected from the range of [1 – 500]. Similarly, in kNN, the nearest neighbors are set to $k = [2 - 20]$, SVM is tested with polynomial kernel function, and ensemble learning classifiers are trained using tree-based model with [1 – 100] number of trees. All such parameters are carefully tuned and optimized during the experimental setup. All these experiments are carried out using MATLAB (2014b) on an Intel Core i5 3.20 GHz CPU with 12 GB of RAM.

5. Experimental Results

In this section, we performed a set of experiments to evaluate our proposed FSAM pipeline using both ROSIS and AVIRIS sensors datasets. Evaluating ROSIS sensor datasets is more challenging classification problem dominated by complex urban classes and nested regions than AVIRIS. Here we evaluate the influence of the number labeled samples on the classification performance achieved by several classifiers. Figures 3 and 4 shows the overall and kappa (κ) accuracy as a function of the number of labeled samples obtained by FSAM, i.e., fuzziness and SAM diversity-based active selection of most informative and diverse samples in each iteration. These labeled samples were selected by machine-machine interaction which significantly reduces the cost in terms of labeled collection through

human supervisor which is the key aspects of automatic AL methods. The plots are shown in Figures 3 and 4 and generated based on only selected samples in contrast to the entire population which reveals clear advantages of using fewer labeled samples for FSAM pipeline.

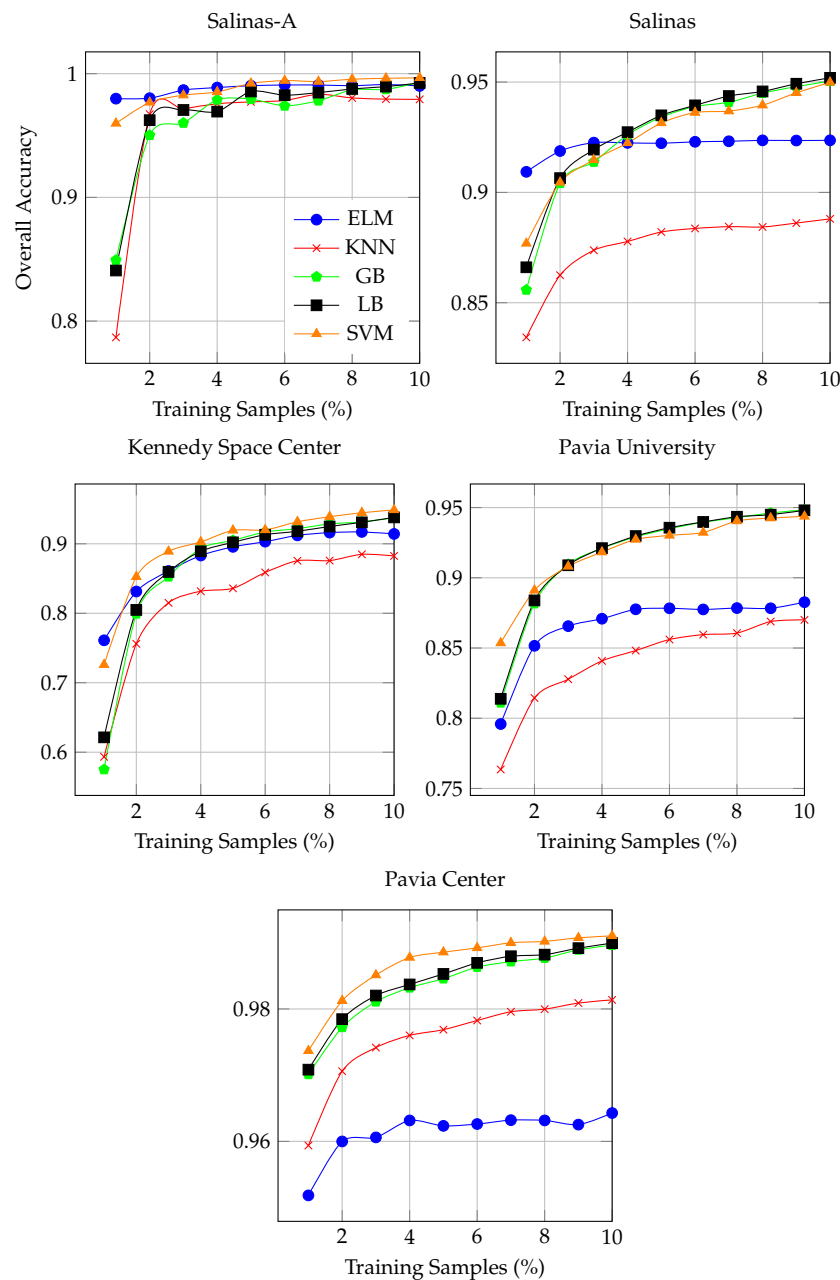


Figure 3. Overall accuracy with different number of training samples (%) selected in each iteration from different datasets. It is perceived from the above figure that by including the samples back to the training set, the classification results are significantly improved for all the classifiers. Moreover, it can be seen that SVM and ELM classifiers are more robust. For examples, with 2% actively selected samples in ELM classifier case, only 2% difference in the classification with a different number of samples can be observed, however, for the KNN and SVM classifiers, the difference is quite high.

From Figures 3 and 4, it can be observed that FSAM greatly improved the accuracy. The results also reveal that SVM and LB outperformed other classifiers in most cases, whereas, as expected, KNN provides lower classification accuracy than SVM and LB, since the candidates are more relevant when the samples are acquired from the class boundaries. Furthermore, it can also be observed that SVM always performed better than KNN, ELM, and ensemble learning classifiers. ELM could perform

better with more number of hidden neurons on more powerful machines. For instance, when the 2% of labeled samples were used, the performance has been significantly increased in contrast to the 1% of actively selected samples. These observations confirm that FSAM can greatly improve the results obtained by different classifiers based on a small portion from the entire population, i.e., the classifiers trained using a limited number of selected labeled samples can produce better generalization performance rather than selecting the bulk amount of label training samples.

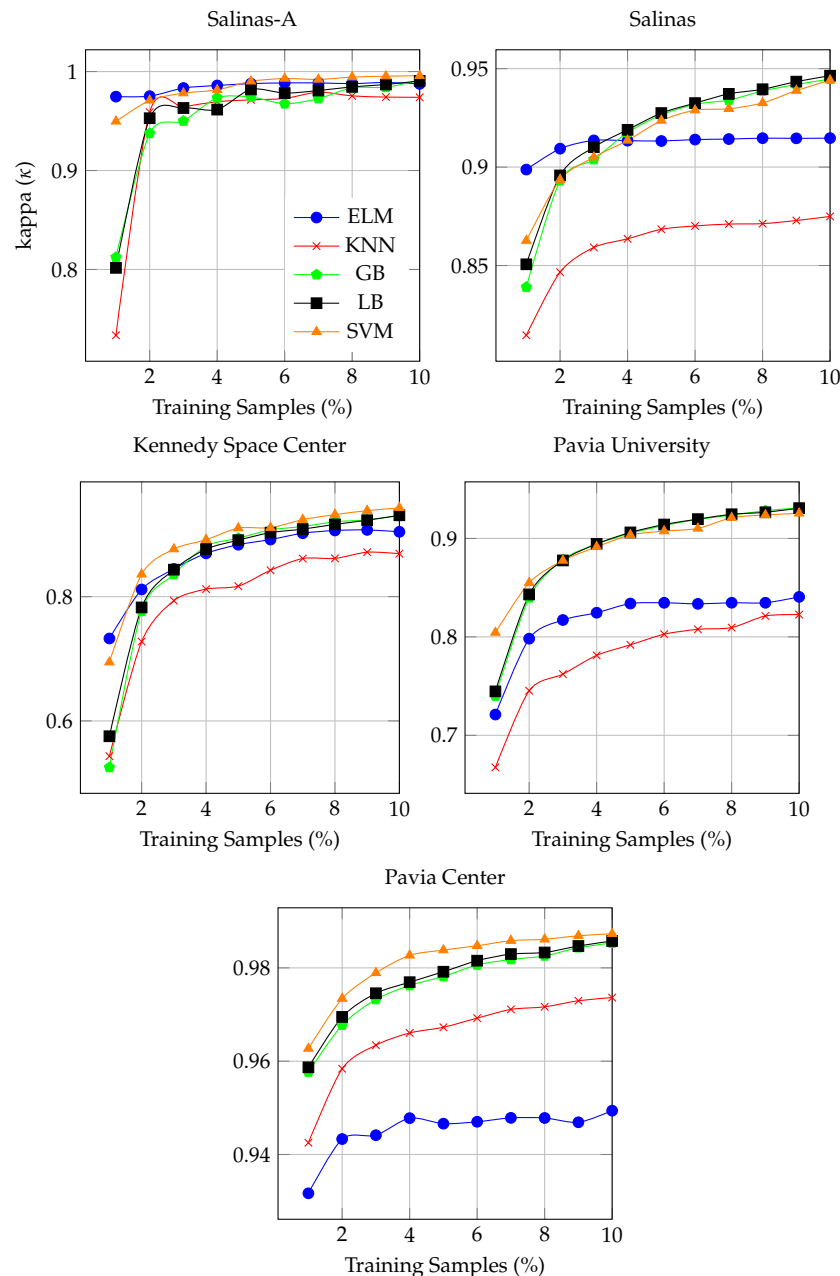


Figure 4. Kappa (κ) accuracy with different number of training samples (%) selected in each iteration from Salinas-A, Salinas, Kennedy Space Center, Pavia University, and Pavia Center datasets respectively. It is perceived from the above figure that by including the samples back to the training set, the classification results in terms of kappa κ are significantly improved for all the classifiers. Moreover, it can be seen that SVM and ELM classifiers are more robust than ensemble and KNN classifiers. For examples, with 2% actively selected samples in ELM classifier case, only 2% difference in the classification with a different number of samples can be observed, however, for the KNN and SVM classifiers, the difference is quite high. Similar observations can be made for ensemble learning models.

It is perceived from Figures 3 and 4 that by including the samples back to the training set, the classification results are significantly improved for all the classifiers. Moreover, it can be seen that SVM and ELM classifiers are more robust than ensemble and KNN classifiers. For examples, with 1% actively selected samples in ELM classifier case, only 2% difference in classification with different number of samples can be observed, however, for the KNN and SVM classifiers, the difference is quite high. Similar observations can be made for ensemble models.

In order to present the classification results in geographical fashion for both ROSIS and AVIRIS sensors datasets, Figures 5–9 shows ground truths segmentation of all experimental datasets used in this work. These ground truths are generated using 2% of actively selected samples by FSAM pipeline. In all the experiments, we provide the quantity of labeled training samples and the test samples which provide an indication of the number of true versus estimated labels used in the experiments. It can be observed from listed results, that our proposed fuzziness and diversity-based active labeled sample selection pipeline is quite robust as it achieved higher classification results which are way better or at least comparable with several state-of-the-art AL methods.

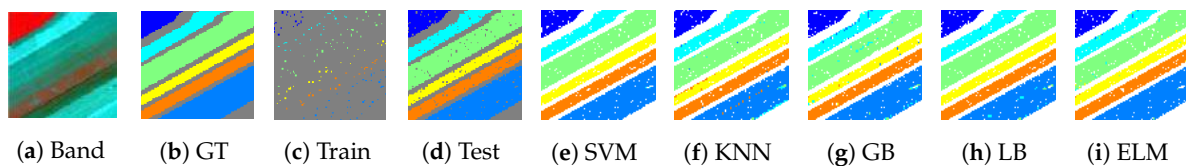


Figure 5. Salinas-A: (a): Ground Band, (b): True Ground Truths, (c): Training Ground Truths, (d): Test Ground Truths, and ground truths predicted by (e): SVM, (f): KNN, (g): GB, (h): LB, and (i): ELM classifier with 2% of selected training samples.

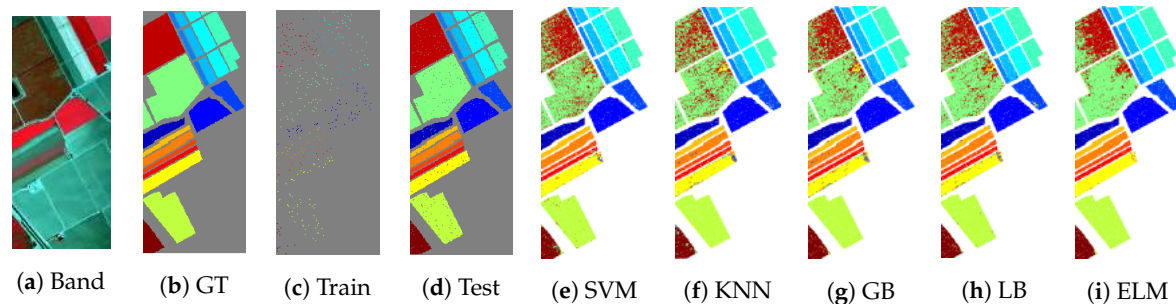


Figure 6. Salinas: (a) Ground Band, (b): True Ground Truths, (c): Training Ground Truths, (d): Test Ground Truths, and ground truths predicted by (e): SVM, (f): KNN, (g): GB, (h): LB, and (i): ELM classifier with 2% of selected training samples.

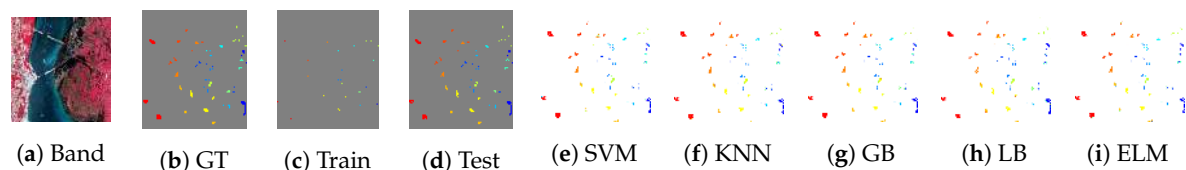


Figure 7. Kennedy Space Center: (a) Ground Band, (b): True Ground Truths, (c): Training Ground Truths, (d): Test Ground Truths, and ground truths predicted by (e): SVM, (f): KNN, (g): GB, (h): LB, and (i): ELM classifiers with 2% of selected training samples.

To better analyze the performance of FSAM on ROSIS and AVIRIS datasets, Table 3 shows the statistical significance in terms of recall, precision, and F1-score tests. The experiments shown in Table 3 are performed with 2% of actively selected labeled samples from each class for all experimental datasets. Table 3 is produced to support the results shown in Figures 3–9 for both AVIRIS and ROSIS sensor datasets. The global recall, precision, and F1-score for each classifier of these results are obtained using 5 Monte Carlo runs. Furthermore, these Tables shows the statistical significance of FSAM in terms of recall, precision, and F1-score with the 99% confidence interval. The obtained values indicate

the ability of FSAM to correctly identify the unseen samples in which each classifier was trained on a very small amount of labeled training samples. For any good model, precision, recall, and F1-score values should be greater than 80% in average, and in our case, these values are almost above 80% for all experimental datasets and for all classifiers, demonstrating that the proposed FSAM-AL pipeline is not classifier sensitive.

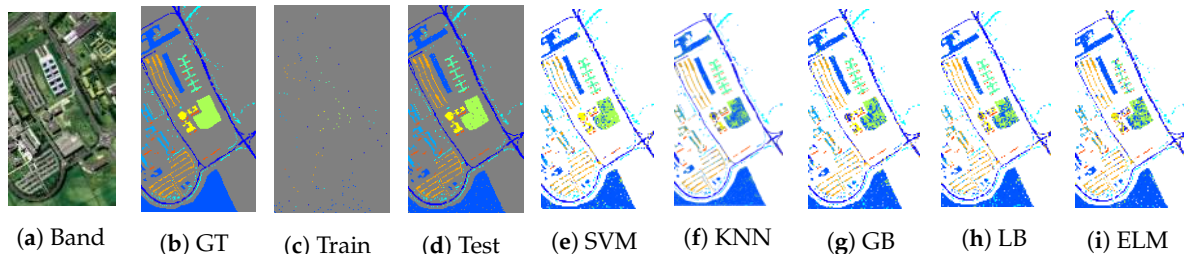


Figure 8. Pavia University: (a) Ground Band, (b): True Ground Truths, (c): Training Ground Truths, (d): Test Ground Truths, and ground truths predicted by (e): SVM, (f): KNN, (g): GB, (h): LB, and (i): ELM classifier with 2% of selected training samples.

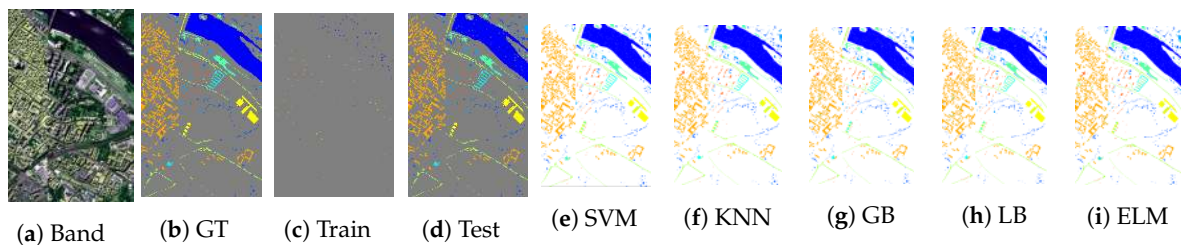


Figure 9. Pavia Center: (a) Ground Band, (b): True Ground Truths, (c): Training Ground Truths, (d): Test Ground Truths, and ground truths predicted by (e): SVM, (f): KNN, (g): GB, (h): LB, and (i): ELM classifier with 2% of selected training samples.

Table 3. Statistical applicability of our proposed FSAM samples selection method. Each classifier is trained with 2% of actively selected training samples.

Tests	ELM	KNN	GB	LB	SVM
Salinas-A Dataset.					
Recall	0.9852 ± 0.0037	0.9489 ± 0.0528	0.9644 ± 0.0281	0.9650 ± 0.0315	0.9855 ± 0.0092
Precision	0.9903 ± 0.0029	0.9567 ± 0.0352	0.9649 ± 0.0275	0.9672 ± 0.0287	0.9885 ± 0.0071
F1 Score	0.9875 ± 0.0031	0.9459 ± 0.0609	0.9639 ± 0.0270	0.9654 ± 0.0300	0.9867 ± 0.0076
Salinas Dataset.					
Recall	0.9544 ± 0.0032	0.9247 ± 0.0162	0.9551 ± 0.0153	0.9580 ± 0.0135	0.9583 ± 0.0115
Precision	0.9584 ± 0.0030	0.9189 ± 0.0117	0.9596 ± 0.0127	0.9603 ± 0.0117	0.9642 ± 0.0091
F1 Score	0.9552 ± 0.0026	0.9225 ± 0.0129	0.9570 ± 0.0138	0.9588 ± 0.0124	0.9611 ± 0.0101
Kennedy Space Center Dataset.					
Recall	0.8220 ± 0.0518	0.7513 ± 0.0739	0.8158 ± 0.0839	0.8159 ± 0.0790	0.8546 ± 0.0579
Precision	0.8445 ± 0.0423	0.8375 ± 0.0638	0.8315 ± 0.0751	0.8344 ± 0.0683	0.8596 ± 0.0509
F1 Score	0.8260 ± 0.0469	0.7491 ± 0.0760	0.8183 ± 0.0815	0.8195 ± 0.0775	0.8542 ± 0.0533
Pavia University Dataset.					
Recall	0.7782 ± 0.0283	0.7954 ± 0.0346	0.8838 ± 0.0391	0.8856 ± 0.0376	0.8858 ± 0.0287
Precision	0.8708 ± 0.0229	0.9508 ± 0.0090	0.9165 ± 0.0247	0.9146 ± 0.0252	0.8929 ± 0.0203
F1 Score	0.8107 ± 0.0255	0.8122 ± 0.0316	0.8973 ± 0.0336	0.8976 ± 0.0327	0.8886 ± 0.0241
Pavia Center Dataset.					
Recall	0.8913 ± 0.0073	0.9281 ± 0.0151	0.9493 ± 0.0124	0.9509 ± 0.0120	0.9568 ± 0.0115
Precision	0.9187 ± 0.0066	0.9125 ± 0.0136	0.9469 ± 0.0112	0.9489 ± 0.0107	0.9572 ± 0.0119
F1 Score	0.8999 ± 0.0064	0.9248 ± 0.0129	0.9480 ± 0.0116	0.9498 ± 0.0112	0.9568 ± 0.0117

6. Comparison and Discussion

The most advanced developments in AL are single pass context and hybrid AL. These techniques combine the concepts of incremental and adaptive learning from the field of online and traditional machine learning. These advancements have resulted in a substantial number of AL methods. The most classical and well studied AL methods include, for example, the works [60,61] focused on online learning. These works specifically designed for on-line single-pass setting in which the data stream samples arrive continuously, thus, does not allow classifier re-training. Furthermore, these works focused on close concepts of conflict and ignorance. Conflict models how close a query point is to the actual class boundary and ignorance represents the distance between already seen training samples and a new sample.

Similar works proposed in [62,63] focused only on early AL strategies such as early-stage experimental design problems. The TED method was proposed to select the samples using robust AL method incorporated with structured sparsity-inducing norms to relax the NP-hard objective of the convex formulation. Thus, these works only focused on selecting an optimal set of initial samples to kick-start the AL. However, the superiority of our proposed FSAM pipeline is that it shows state-of-the-art performance independent of how the initial labeled training samples are selected. Such methods can easily be integrated into the works which utilize the decision boundary based sample selection methods.

A novel tri-training semi-supervised hyperspectral image classification method based on regularized local discriminant embedding feature extraction (RLDE) was proposed in [64]. In this work, the RLDE process is used for optimal number of feature extraction to overcome the limitation of singular values and over-fitting of local Fisher discriminant analysis and local discriminant embedding. At a later stage, active learning method is used to select the informative samples from the candidate set. This work solves the singularity issues of LDA, however, this may include the redundant samples back to the training set which do not provide any new information to the classifier.

Spatial-spectral multiview 3D Gabor inspired active learning for hyperspectral image classification method was proposed in [65]. Trivial multiview active learning methods can make a comprehensive analysis of both sample selection and object characterization in active learning by using several features of multiple views. However, multiview cannot effectively exploit spatial-spectral information by respecting the 3D nature of hyperspectral imaging, therefore, the sample selection method in multiview is only based on the disagreement of multiple views. To overcome such problems, J. Hu, et al. [65] proposed a two-step 3D Gabor inspired multiview method for hyperspectral image classification. The first step consists of the view generation step, in which a 3D Gabor filter was used to generate multiple cubes with limited bands and utilize the features assessment strategies to select cubes for constructing views. On a second stage, an active learning method was presented which used both external and internal uncertainty estimation of views. More specifically, posterior probability distribution was used to learn the internal uncertainty of each individual independent view and external uncertainty was computed using inconsistency between the views.

Of course, the frameworks proposed in the above papers can be easily integrated with our proposed FSAM sample selection method instead of selecting the samples based on uncertainty or tri-training methods. We initialize our active learning method from 100 number of randomly selected labeled training samples and we experimentally demonstrate that randomly increasing the size of the training set slightly increases the accuracy nevertheless the classifiers become computationally complex. Therefore, at the first step, we decided to separate the set of misclassified samples which have higher fuzziness values (samples fuzziness magnitude between 0.7–1.0). We then select a specific percentage of misclassified samples which have higher fuzziness to compute the spectral angle among the reference training samples. We then fused a specific percentage of selected samples with the original training set to retrain the classifier from scratch for better generalization and classification performance on those samples which were initially misclassified by the same classifier.

More specifically, the proposed solution has been rigorously investigated through comparison against some significant works recently published in the HSI classification area, adopting different sample selection methods such as random sampling (RS), mutual information (MI), breaking ties (BT), modified breaking ties (MBT), uncertainty, and fuzziness as introduced in Section 4. This comparison is based on the Botswana hyperspectral dataset acquired by the NASA EO-1 Satellite Hyperion sensor [58,66]. The experiments are based on five Monte Carlo runs with 100 initial training samples selected from this dataset. In each iteration, the training set size has been increased of 50 samples selected by a specific method among the ones to be compared. The results thus obtained are presented in the Tables 4–8. Based on such results, we can argue that the FSAM pipeline outperforms the other solutions taken into account in these experiments. This is due to the dual soft thresholding method for selection of the most informative as well as spatially heterogeneous labeled training samples. Furthermore, another benefit of the proposed FSAM solution is that it systematically selects the most informative but least redundant labeled training samples by machine–machine interaction without involving any supervisor, automatically, while the other AL frameworks need that a supervisor selects the samples at each iteration, manually.

Table 4. Kappa (κ) accuracy obtained by SVM Classifier with different number of training samples selected in each iteration from Botswana dataset with different sample selection methods from literature.

Sample Selection Method	Number of Training Samples									
	50	100	150	200	250	300	350	400	450	500
	Kappa Accuracy									
Random Sampling [28]	0.8156	0.8483	0.8738	0.8886	0.9005	0.9101	0.9170	0.9151	0.9163	0.9221
Mutual Information [30–32]	0.8149	0.8437	0.8602	0.8798	0.8863	0.9002	0.9108	0.9217	0.9195	0.9302
Breaking Ties [33]	0.8163	0.8316	0.8401	0.8561	0.8778	0.8919	0.9008	0.9014	0.9087	0.9128
Modified Breaking Ties [34,35]	0.8156	0.8522	0.8563	0.8893	0.9007	0.9040	0.9068	0.9136	0.9138	0.9103
Fuzziness [11]	0.8174	0.8129	0.8422	0.8648	0.8755	0.8934	0.8989	0.8986	0.9156	0.9119
FSAM	0.8167	0.8749	0.9027	0.9091	0.9493	0.9556	0.9668	0.9788	0.9928	0.9984

Table 5. Kappa (κ) accuracy obtained by ELM Classifier with different number of training samples selected in each iteration from Botswana dataset with different sample selection methods from literature.

Sample Selection Method	Number of Training Samples									
	50	100	150	200	250	300	350	400	450	500
	Kappa Accuracy									
Random Sampling [28]	0.8094	0.8253	0.8364	0.8564	0.8648	0.8730	0.8919	0.8958	0.9063	0.9140
Mutual Information [30–32]	0.8051	0.8246	0.8430	0.8538	0.8699	0.8772	0.8881	0.8962	0.8983	0.9070
Breaking Ties [33]	0.8051	0.8174	0.8392	0.8607	0.8680	0.8744	0.8819	0.8963	0.8927	0.9022
Modified Breaking Ties [34,35]	0.7961	0.8216	0.8563	0.8654	0.8718	0.8769	0.8896	0.9012	0.9081	0.9149
Fuzziness [11]	0.7958	0.8224	0.8463	0.8513	0.8606	0.8733	0.8780	0.8841	0.9008	0.9083
FSAM	0.8021	0.8385	0.8544	0.8846	0.8923	0.8968	0.9213	0.9355	0.9492	0.9551

Table 6. Kappa (κ) accuracy obtained by KNN Classifier with different number of training samples selected in each iteration from Botswana dataset with different sample selection methods from literature.

Sample Selection Method	Number of Training Samples									
	50	100	150	200	250	300	350	400	450	500
	Kappa Accuracy									
Random Sampling [28]	0.7854	0.8145	0.8158	0.8428	0.8547	0.8556	0.8603	0.8640	0.8695	0.8757
Mutual Information [30–32]	0.7854	0.8029	0.8154	0.8342	0.8485	0.8519	0.8592	0.8653	0.8727	0.8814
Breaking Ties [33]	0.7854	0.8205	0.8330	0.8466	0.8469	0.8541	0.8626	0.8731	0.8760	0.8757
Modified Breaking Ties [34,35]	0.7854	0.8396	0.8463	0.8474	0.8584	0.8635	0.8685	0.8749	0.8813	0.8841
Fuzziness [11]	0.7854	0.8248	0.8298	0.8445	0.8529	0.8584	0.8628	0.8678	0.8702	0.8771
FSAM	0.7854	0.8369	0.8512	0.8842	0.8919	0.9189	0.9236	0.9469	0.9584	0.9625

Table 7. Kappa (κ) accuracy obtained by **GB Classifier** with different number of training samples selected in each iteration from Botswana dataset with different sample selection methods from literature.

Sample Selection Method	Number of Training Samples									
	50	100	150	200	250	300	350	400	450	500
	Kappa Accuracy									
Random Sampling [28]	0.8140	0.8272	0.8342	0.8358	0.8479	0.8597	0.8649	0.8703	0.8675	0.8717
Mutual Information [30–32]	0.8139	0.7941	0.8118	0.8406	0.8620	0.8608	0.8687	0.8750	0.8787	0.8712
Breaking Ties [33]	0.8139	0.7875	0.8318	0.8355	0.8625	0.8691	0.8795	0.8895	0.8935	0.8928
Modified Breaking Ties [34,35]	0.8139	0.8067	0.8404	0.852	0.8570	0.8612	0.8711	0.8772	0.8750	0.8795
Fuzziness [11]	0.8139	0.8470	0.8488	0.8524	0.8559	0.8658	0.8692	0.8755	0.8782	0.8853
FSAM	0.8140	0.8054	0.8605	0.8852	0.9060	0.9268	0.9247	0.9280	0.9455	0.9592

Table 8. Kappa (κ) accuracy obtained by **LB Classifier** with different number of training samples selected in each iteration from Botswana dataset with different sample selection methods from literature.

Sample Selection Method	Number of Training Samples									
	50	100	150	200	250	300	350	400	450	500
	Kappa Accuracy									
Random Sampling [28]	0.8074	0.8105	0.8275	0.8359	0.8419	0.8429	0.8482	0.8620	0.8699	0.8768
Mutual Information [30–32]	0.8073	0.8212	0.83151	0.8366	0.8409	0.8442	0.8564	0.854	0.8571	0.8565
Breaking Ties [33]	0.8074	0.8155	0.8300	0.8293	0.8337	0.8467	0.8513	0.8490	0.8582	0.8679
Modified Breaking Ties [34,35]	0.8074	0.8303	0.8402	0.8422	0.8484	0.8649	0.8704	0.8813	0.8816	0.8794
Fuzziness [11]	0.8073	0.8126	0.8236	0.8411	0.8447	0.8586	0.8626	0.8684	0.8669	0.8709
FSAM	0.8073	0.8118	0.8226	0.8779	0.8993	0.9094	0.9216	0.9402	0.9475	0.9588

By the Botswana dataset we experimentally demonstrated that FSAM outperforms all other sample selection methods, i.e., random selection, mutual information, breaking ties, modified breaking ties, and fuzziness in terms of accuracy, starting from the same classifiers and the same number of labeled training samples as shown in Tables 4–8. Furthermore, all these sample selection methods are more often subjective and tends to bring redundancy into the classifiers. reducing the generalization performance of the classifiers. More specifically, the number of samples required to learn a model in FSAM can be much lower than the number of selected samples. In such scenarios, there is a risk, however, that the learning model may get overwhelmed because of the uninformative or spatially miscellaneous samples selected by query function.

7. Conclusions

The classification of multiclass spatial–spectral HSI with a small labeled training sample size is a challenging task. To overcome this problem, this paper introduces a customized AL pipeline for HSI to reduce the sample selection bias while maintaining the data stability in the spatial domain.

The proposed FSAM pipeline differs from traditional AL methods in three relevant aspects. First, instead of simply using the uncertainty of samples to select new samples, it utilizes the fuzziness measure associated with the confidence of the training model in classifying those samples correctly. Second, it couples the samples' fuzziness with their diversity to select new training samples which simultaneously minimize the error among the training samples while maximizing the spectral angle between the selected sample and the existing training samples. In our work, instead of measuring angle-based distances among all new samples and all existing training samples, a reference sample is selected from within the training set against which the diversity of the new samples is measured. This achieves the same goal while reducing the computational overhead as the size of training set is always much smaller than the validation set which is the source of new samples. Thirdly, the FSAM keeps the pool of new samples balanced, giving equal representation to all classes, which is achieved via softening the thresholds at run time.

Experimental results on five benchmark datasets demonstrate that the proposed FSAM leads to an increased predictive power regarding kappa (κ) and overall accuracy, precision, recall, and F1-Score parameters. A comparison of FSAM with state-of-the-art sample selection method is performed, confirming that the FSAM is effective in terms of overall accuracy and κ , also with few training samples.

However, the main drawback of SAM is spectral mixture problems, i.e., SAM assumes that the reference spectra chosen to classify the HSI represents the pure spectra. Such problem occurs when the HSI is in low or medium spatial resolution. Furthermore, as we know, the surface of the earth is widely heterogeneous and complex, thus containing many mixed samples. The spectral confusion in samples can lead to overestimation or underestimation errors for a spectral signatures. Our future research direction aims to address such limitations to classify low or mid spatial resolution hyperspectral images in a computationally efficient way. Further work will be directed toward testing the FSAM pipeline in different analysis scenarios dominated by the limited availability of training samples a priori.

Author Contributions: Conceptualization, M.A.; methodology, M.A.; validation, M.A.; formal analysis, M.A., A.M.K.; investigation, M.A.; writing—original draft preparation, M.A.; writing—review and editing, M.A., A.K., A.M.K., M.M., S.D., A.S. and O.N.; visualization, M.A.; supervision, A.M.K.

Funding: This research work was partially funded by Innopolis University.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

HSI	Hyperspectral Imaging
AL	Active Learning
SAM	Spectral Angle Mapper
FSAM	Fuzziness and Spectral Angle Mapper (SAM)-based Active Sample Selection
D_T	Training Samples
D_V	Test Samples
D_V^H	High Fuzziness Samples from Test Set
SVM	Support Vector Machine
KNN	K Nearest Neighbours
LB	Logistic Boost
ELM	Extreme Learning Machine
QBC	Query by Committee
RS	Random Sampling
MI	Mutual Information
BT	Breaking Ties
MBT	Modified Breaking Ties
κ	kappa
OA	Overall Accuracy

References

1. Schneider, A.; Feussner, H. *Diagnostic Procedures*; Institute of Minimally Invasive Interdisciplinary Therapeutic Interventions (MITI), Technische Universität München (TUM), Biomedical Engineering in Gastrointestinal Surgery: London, UK, 2017; Chapter 5.
2. Ahmad, M.; Alqarni, M.; Khan, A.M.; Hussain, R.; Mazzara, M.; Distefano, S. Segmented and non-segmented stacked denoising autoencoder for hyperspectral band reduction. *Optik Int. J. Light Electron Opt.* **2019**, *180*, 370–378. [[CrossRef](#)]
3. Qu, Y.; Qi, H.; Kwan, C. Unsupervised Sparse Dirichlet-Net for Hyperspectral Image Super-Resolution. *arXiv* **2018**, arXiv:1804.05042.
4. Ahmad, M.; Bashir, A.K.; Khan, A.M. Metric similarity regularizer to enhance pixel similarity performance for hyperspectral unmixing. *Optik Int. J. Light Electron Opt.* **2017**, *140*, 86–95. [[CrossRef](#)]

5. Ahmad, M.; Khan, A.M.; Mazzara, M.; Distefano, S. Multi-layer Extreme Learning Machine-based Autoencoder for Hyperspectral Image Classification. In Proceedings of the 14th International Conference on Computer Vision Theory and Applications (VISAPP'19), Prague, Czech Republic, 25–27 February 2019.
6. Li, J.; Bioucas-Dias, J.M.; Plaza, A. Spectral-Spatial Hyperspectral Image Segmentation Using Subspace Multinomial Logistic Regression and Markov Random Fields. *IEEE Trans. Geosci. Remote Sens.* **2012**, *50*, 809–823. [[CrossRef](#)]
7. Xia, J.; Du, P.; He, X.; Chanussot, J. Hyperspectral Remote Sensing Image Classification Based on Rotation Forest. *IEEE Geosci. Remote Sens. Lett.* **2014**, *11*, 239–243. [[CrossRef](#)]
8. Pan, B.; Shi, Z.; Xu, X. Hierarchical Guidance Filtering-Based Ensemble Classification for Hyperspectral Images. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 4177–4189. [[CrossRef](#)]
9. Pan, B.; Shi, Z.; Xia, X. R-VCANet: A New Deep-Learning-Based Hyperspectral Image Classification Method. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2017**, *10*, 1975–1986. [[CrossRef](#)]
10. Tan, K.; Du, P. Hyperspectral Remote Sensing Image Classification Based on Support Vector Machine. *J. Infrared Millim. Waves* **2013**, *27*, 123–128. [[CrossRef](#)]
11. Ahmad, M.; Protasov, S.; Khan, A.M.; Hussain, R.; Khattak, A.M.; Khan, W.A. Fuzziness-based active learning framework to enhance hyperspectral image classification performance for discriminative and generative classifiers. *PLoS ONE* **2018**, *13*, e0188996. [[CrossRef](#)]
12. Hughes, G. On the mean accuracy of statistical pattern recognizers. *IEEE Trans. Inf. Theory* **1968**, *14*, 55–63. [[CrossRef](#)]
13. Persello, C.; Bruzzone, L. Active and Semisupervised Learning for the Classification of Remote Sensing Images. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 6937–6956. [[CrossRef](#)]
14. Yang, L.; Yang, S.; Jin, P.; Zhang, R. Semi-Supervised Hyperspectral Image Classification Using Spatio-Spectral Laplacian Support Vector Machine. *IEEE Geosci. Remote Sens. Lett.* **2014**, *11*, 651–655. [[CrossRef](#)]
15. Zhou, Z.H.; Li, M. Tri-Training: Exploiting Unlabeled Data Using Three Classifiers. *IEEE Trans. Knowl. Data Eng.* **2005**, *17*, 1529–1541. [[CrossRef](#)]
16. Ahmad, M.; Khan, A.M.; Hussain, R. Graph-based spatial-spectral feature learning for hyperspectral image classification. *IET Image Process.* **2017**, *11*, 1310–1316. [[CrossRef](#)]
17. Ly, N.H.; Qian, D.; Fowler, J.E. Sparse Graph-Based Discriminant Analysis for Hyperspectral Imagery. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 3872–3884. [[CrossRef](#)]
18. Zhang, L.; Du, B. Recent advances in hyperspectral image processing. *Geo-Spat. Inf. Sci.* **2012**, *15*, 143–156. [[CrossRef](#)]
19. Ahmad, M.; Khan, A.M.; Hussain, R.; Protasov, S.; Chow, F.; Khattak, A.M. Unsupervised geometrical feature learning from hyperspectral data. In Proceedings of the 2016 IEEE Symposium Series on Computational Intelligence (SSCI), Athens, Greece, 6–9 December 2016; pp. 1–6.
20. Tuia, D.; Ratle, F.; Pacifici, F.; Kanevski, M.F.; Emery, W.J. Active Learning Methods for Remote Sensing Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2009**, *47*, 2218–2232. [[CrossRef](#)]
21. Dopido, I.; Jun, L.; Marpu, P.R.; Plaza, J.M.A.; Dias, B.; Benediktsson, J.A. Semisupervised Self-Learning for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2013**, *51*, 4032–4044. [[CrossRef](#)]
22. Yang, L.; MacEachren, A.M.; Mitra, P.; Onorati, T. Visually-Enabled Active Deep Learning for (Geo) Text and Image Classification: A Review. *ISPRS Int. J. Geo-Inf.* **2018**, *7*, 65. [[CrossRef](#)]
23. Yu, H.; Yang, X.; Zheng, S.; Sun, C. Active Learning from Imbalanced Data: A Solution of Online Weighted Extreme Learning Machine. *IEEE Trans. Neural Netw. Learn. Syst.* **2018**, *30*, 1088–1103. [[CrossRef](#)]
24. Liu, C.; He, L.; Li, Z.; Li, J. Feature-Driven Active Learning for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 341–354. [[CrossRef](#)]
25. Mountrakis, G.; Im, J.; Ogole, C. Support vector machines in remote sensing: A review. *ISPRS J. Photogramm. Remote Sens.* **2011**, *66*, 247–259. [[CrossRef](#)]
26. Pasolli, E.; Melgani, F.; Tuia, D.; Pacifici, F.; Emery, W.J. Improving active learning methods using spatial information. In Proceedings of the 2011 IEEE International Geoscience and Remote Sensing Symposium, Sendai, Japan, 1–5 August 2011; pp. 3923–3926.
27. Liu, A.; Jun, G.; Ghosh, J. Active learning of hyperspectral data with spatially dependent label acquisition costs. In Proceedings of the 2009 IEEE International Geoscience and Remote Sensing Symposium, Cape Town, South Africa, 12–17 July 2009.

28. Tuia, D.; Volpi, M.; Copa, L.; Kanevski, M.; Munoz-Mari, J. A Survey of Active Learning Algorithms for Supervised Remote Sensing Image Classification. *IEEE J. Sel. Top. Signal Process.* **2011**, *5*, 606–617. [[CrossRef](#)]
29. Pasolli, E.; Melgani, F.; Tuia, D.; Pacifici, F.; Emery, W.J. SVM Active Learning Approach for Image Classification Using Spatial Information. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 2217–2233. [[CrossRef](#)]
30. MacKay, D.J.C. Information-Based Objective Functions for Active Data Selection. *Neural Comput.* **1992**, *4*, 590–604. [[CrossRef](#)]
31. Krishnapuram, B.; Williams, D.; Xue, Y.; Carin, L.; Figueiredo, M.; Hartemink, A.J. On Semi-Supervised Classification. In *Advances in Neural Information Processing Systems 17*; Saul, L.K., Weiss, Y., Bottou, L., Eds.; MIT Press: Cambridge, MA, USA, 2005; pp. 721–728.
32. Li, J.; Bioucas-Dias, J.M.; Plaza, A. Spectral-Spatial Classification of Hyperspectral Data Using Loopy Belief Propagation and Active Learning. *IEEE Trans. Geosci. Remote Sens.* **2013**, *51*, 844–856. [[CrossRef](#)]
33. Luo, T.; Kramer, K.; Samson, S.; Remsen, A.; Goldgof, D.B.; Hall, L.O.; Hopkins, T. Active learning to recognize multiple types of plankton. In Proceedings of the 17th International Conference on Pattern Recognition, Cambridge, UK, 23–26 August 2004; pp. 478–481.
34. Li, J.; Bioucas-Dias, J.M.; Plaza, A. Hyperspectral Image Segmentation Using a New Bayesian Approach With Active Learning. *IEEE Trans. Geosci. Remote Sens.* **2011**, *49*, 3947–3960. [[CrossRef](#)]
35. Shi, Q.; Du, B.; Zhang, L. Spatial Coherence-Based Batch-Mode Active Learning for Remote Sensing Image Classification. *IEEE Trans. Image Process.* **2015**, *24*, 2037–2050. [[CrossRef](#)] [[PubMed](#)]
36. Demir, B.; Persello, C.; Bruzzone, L. Batch-Mode Active-Learning Methods for the Interactive Classification of Remote Sensing Images. *IEEE Trans. Geosci. Remote Sens.* **2011**, *49*, 1014–1031. [[CrossRef](#)]
37. Lewis, D.D.; Gale, A.W. A Sequential Algorithm for Training Text Classifiers. In Proceedings of the 17th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Dublin, Ireland, 3–6 July 1994; pp. 3–12.
38. Di, W.; Crawford, M.M. Active Learning via Multi-View and Local Proximity Co-Regularization for Hyperspectral Image Classification. *IEEE J. Sel. Top. Signal Process.* **2011**, *5*, 618–628. [[CrossRef](#)]
39. Patra, S.; Bhardwaj, K.; Bruzzone, L. A Spectral-Spatial Multicriteria Active Learning Technique for Hyperspectral Image Classification. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2017**, *10*, 5213–5227. [[CrossRef](#)]
40. Li, J. Active learning for hyperspectral image classification with a stacked autoencoders based neural network. In Proceedings of the 2015 7th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), Tokyo, Japan, 2–5 June 2015; pp. 1–4.
41. David, A.C.; Ghahramani, Z. Active Learning with Statistical Models. *J. Artif. Intell. Res.* **1996**, *4*, 129–145.
42. Rajan, S.; Ghosh, J.; Crawford, M.M. An Active Learning Approach to Hyperspectral Data Classification. *IEEE Trans. Geosci. Remote Sens.* **2008**, *46*, 1231–1242. [[CrossRef](#)]
43. Seung, H.S.; Opper, M.; Sompolinsky, H. Query by Committee. In Proceedings of the Fifth Annual Workshop on Computational Learning Theory, Pittsburgh, PA, USA, 27–29 July 1992; pp. 287–294.
44. Haines, T.; Xiang, T. Active Learning using Dirichlet Processes for Rare Class Discovery and Classification. In Proceedings of the British Machine Vision Conference, Dundee, UK, 29 August–2 September 2011; pp. 9.1–9.11.
45. Michel, J.; Malik, J.; Inglada, J. Lazy yet efficient land-cover map generation for HR optical images. In Proceedings of the 2010 IEEE International Geoscience and Remote Sensing Symposium, Honolulu, HI, USA, 25–30 July 2010; pp. 1863–1866.
46. Borisov, A.; Tuv, E.; Runger, G. Active Batch Learning with Stochastic Query-by-Forest (SQBF). *Proc. Mach. Learn. Res.* **2011**, *16*, 59–69.
47. Munoz-Mari, J.; Tuia, D.; Camps-Valls, G. Semisupervised Classification of Remote Sensing Images with Active Queries. *IEEE Trans. Geosci. Remote Sens.* **2012**, *50*, 3751–3763. [[CrossRef](#)]
48. He, L.; Li, J.; Liu, C.; Li, S. Recent Advances on Spectral-Spatial Hyperspectral Image Classification: An Overview and New Guidelines. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 1579–1597. [[CrossRef](#)]
49. Yu, H.; Sun, C.; Yang, W.; Yang, X.; Zuo, X. AL-ELM: One uncertainty-based active learning algorithm using extreme learning machine. *Neurocomputing* **2015**, *166*, 140–150. [[CrossRef](#)]
50. Kumar, S.; Hebert, M. Discriminative Random Fields. *Int. J. Comput. Vis.* **2006**, *68*, 179–201. [[CrossRef](#)]
51. Luca, A.D.; Termini, S. A Definition of a Non-Probabilistic Entropy in the Setting of Fuzzy Sets Theory. *J. Inf. Control* **1972**, *20*, 301–312. [[CrossRef](#)]

52. Yeung, D.S.; Trillas, E. Measures of Fuzziness under Different Uses of Fuzzy Sets. In *Advances in Computational Intelligence*; Springer: Berlin/Heidelberg, Germany, 2012; pp. 25–34.
53. Yuhas, R.H.; Goetz, A.F.H.; Boardman, J.W. Discrimination Among Semi-Arid Landscape Endmembers Using the Spectral Angle Mapper (SAM) Algorithm. In *Summaries of the 4th JPL Airborne Earth Science Workshop*; JPL Publication, Summaries of the Third Annual JPL Airborne Geoscience Workshop; NASA: Washington, DC, USA, 1992; pp. 147–149.
54. van der Meer, F.D.; Vazquez-Torres, M.; van Dijk, P.M. Spectral characterization of ophiolite lithologies in the Troodos ophiolite complex of Cyprus and its potential in prospecting for massive sulphide deposits. *Int. J. Remote Sens.* **1997**, *18*, 1245–1257. [[CrossRef](#)]
55. Carvalho, O.A.D.; Meneses, P.R. Spectral Correlation Mapper (SCM); An Improvement on the Spectral Angle Mapper (SAM). In *Summaries of the 9th JPL Airborne Earth Science Workshop*; JPL Publication 00-18; NASA: Washington, DC, USA, 2000.
56. Remondino, F.; Rizzi, A.; Barazzetti, L.; Scaioni, M.; Francesco, F.; Raffaella, B.; Anna, P. Review of Geometric and Radiometric Analyses of Paintings. *Photogramm. Rec.* **2011**, *26*, 439–461. [[CrossRef](#)]
57. Singh, R.S. Evaluation of EO-1 Hyperion Data for Crop Studies in Part of Indo-Gangatic Plains: A Case Study of Meerut District. *Adv. Remote Sens.* **2015**, *4*, 263–269, doi:10.4236/ars.2015.44021. [[CrossRef](#)]
58. Hyperspectral Datasets Description. Available online: http://www.ehu.es/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes (accessed on 30 June 2018).
59. Yang, J.; Yu, H.; Yang, X.; Zuo, X. Imbalanced Extreme Learning Machine Based on Probability Density Estimation. In *Multi-Disciplinary Trends in Artificial Intelligence*; Bikakis, A., Zheng, X., Eds.; Springer International Publishing: Cham, Switzerland, 2015; pp. 160–167.
60. Woodward, M.; Finn, C. Active One-shot Learning. *arXiv* **2017**, arXiv:1702.06559.
61. Lughofer, E. Single-pass active learning with conflict and ignorance. *Evol. Syst.* **2012**, *3*, 251–271. [[CrossRef](#)]
62. Liu, W.; Chang, X.; Chen, L.; Yang, Y. Early Active Learning with Pairwise Constraint for Person Re-identification. In *Machine Learning and Knowledge Discovery in Databases*; Ceci, M., Hollmén, J., Todorovski, L., Vens, C., Džeroski, S., Eds.; Springer International Publishing: Cham, Switzerland, 2017; pp. 103–118.
63. Nie, F.; Wang, H.; Huang, H.; Ding, C. Early Active Learning via Robust Representation and Structured Sparsity. In *Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence, Beijing, China, 3–9 August 2013*; pp. 1572–1578.
64. Ou, D.; Tan, K.; Du, Q.; Zhu, J.; Wang, X.; Chen, Y. A Novel Tri-Training Technique for the Semi-Supervised Classification of Hyperspectral Images Based on Regularized Local Discriminant Embedding Feature Extraction. *Remote Sens.* **2019**, *11*, 654. [[CrossRef](#)]
65. Hu, J.; He, Z.; Li, J.; He, L.; Wang, Y. 3D-Gabor Inspired Multiview Active Learning for Spectral-Spatial Hyperspectral Image Classification. *Remote Sens.* **2018**, *10*, 1070. [[CrossRef](#)]
66. Preet, P.; Batra, S.S.; Jayadeva. Feature Selection for classification of hyperspectral data by minimizing a tight bound on the VC dimension. *arXiv* **2015**, arXiv:1509.08112 [[CrossRef](#)]

