Active Learning for Hyperspectral Image Classification Using Kernel Sparse Representation Classifiers

Amos Bortiew, Swarnajyoti Patra, *Member, IEEE*, and Lorenzo Bruzzone, *Fellow, IEEE*

***Abstract*— Active learning (AL) is one of the popular approaches that can mitigate some of the drawbacks of supervised classification. Although sparse representation classifier (SRC) has already proven to be a robust classifier and successfully used in many applications, it is seldom used jointly with AL. In this letter, we propose a novel AL technique for SRCs. In the proposed model, the query function is designed by combining uncertainty and diversity criteria, both of which are defined by using the SRC in kernel space. The proposed technique outperforms other state-of-the-art methods in terms of classification performance.**

***Index Terms*— Active learning (AL), hyperpsectral image, kernel space, query function, sparse representation.**

I. Introduction

SUPERVISED methods require adequate labeled samples for training. Since labeling requires either field survey or photointerpretation by the experts, it is expensive and time-consuming. To address this issue, semi-supervised learning and active learning (AL) techniques have been developed. Semi-supervised learning exploits unlabeled samples to select certain samples for improving the reliability and the generalization in the training of a classifier [1]. Whereas, AL selects uncertain samples from the unlabeled pool and assigns them a label with the help of a supervisor for updating the training set of the classifier [2], [3]. The idea behind it is that only informative samples need to be labeled for classification. This reduces the labeling cost and the redundancy in the training set without compromising the classification performance.

The fundamental step of AL is to define the query function that selects informative samples from the unlabeled pool for labeling. In batch mode AL, the query function is usually defined by combining uncertainty and diversity criteria. The uncertainty criteria are mostly supervised in nature; i.e., they are defined by exploiting the classifiers. Most of the AL techniques exploit support vector machines (SVMs) classifier for defining uncertainty criterion because of its classification rule characterized by a small set of support vectors (SVs). Some popular and widely used uncertainty criteria are marginal sampling (MS) [4], [5], multiclass-level uncertainty (MCLU) [6], entropy query-by-bagging (EQB) [5], cluster assumption-based histogram thresholding (CAHT) [7], and breaking tie (BT) [8]. Note that the BT criterion is conceptually the same as MCLU criterion except that the first is more generalized (i.e., it can be defined by exploiting any classifiers), whereas the latter is defined by exploiting the one-against-all SVM classifier.

In batch mode, the uncertain samples selected by applying the above AL criterion may have redundancy. In order to select informative (i.e., uncertain and nonredundant) samples, both an uncertainty and a diversity criterion are applied. In [9], angular-based diversity (ABD) is applied to select diverse samples from the selected uncertain samples by maximizing their angle in the spectral domain. The closest SV (cSV) is another diversity defined using the SVM classifier [5]. It selects the subset of uncertain samples, which are closest to distinct SVs. The clustering-based diversity is widely used in AL. K-means, kernel k-means, and self-organizing map (SOM) are the clustering techniques used to select diverse samples [6], [10]. In recent years, other advanced AL techniques have been proposed in the literature [11], [12], [13], [14]. However, the query functions of most of these techniques still rely on the abovementioned criteria. Therefore, improvement of these criteria (especially of the uncertainty criterion) by exploiting a suitable classifier is the main motivation of this research.

Sparse representation classifier (SRC) has proven to be a robust classifier and has successfully been applied to hyperspectral image (HSI) classification [15], [16]. However, a large training set is required for it to produce high classification accuracy, which is difficult to obtain in real-world scenarios. AL will, therefore, address this issue by selecting only informative samples. Despite recent significant advancements, it has been found that SRC is rarely applied in the context of AL. Huo et al. [17] used SRC to define the BT uncertainty criterion for HSI classification but failed to get convincing results. Wang et al. [18] applied sparse modeling to select diverse uncertain samples, but they relied on the SVM-based uncertainty criterion. In this letter, we propose an AL query function that includes uncertainty and diversity criteria, both of which are defined by exploiting the kernel SRC (KSRC).

Manuscript received 11 November 2022; revised 19 February 2023; accepted 29 March 2023. Date of publication 3 April 2023; date of current version 14 April 2023. This work was supported by the Science and Engineering Research Board (SERB) through Government of India under Grant CRG/2020/003018. *(Corresponding author: Swarnajyoti Patra.)*

Amos Bortiew and Swarnajyoti Patra are with the Department of Computer Science and Engineering, Tezpur University, Tezpur, Assam 784028, India (e-mail: swpatra@gmail.com).

Lorenzo Bruzzone is with the Department of Information Engineering and Computer Science, University of Trento, I-38123 Trento, Italy (e-mail: lorenzo.bruzzone@unitn.it).

Digital Object Identifier 10.1109/LGRS.2023.3264283

The main rationale of the proposed approach is to define an effective AL criterion for KSRC. This is important to merge the robustness of this classifier in the analysis of hyperspectral data with the capability to perform the learning phase with a minimum number of training samples selected ad hoc for satisfying the requirements of KSRC. The main contributions of our technique are as follows: 1) it proposed a robust uncertainty criterion by exploiting KSRC and 2) it also proposed a novel diversity criterion by exploiting the dictionary of the KSRC.

The rest of this letter is organized as follows. Section II presents SRCs in brief. Section III describes the proposed AL method in detail. Section IV illustrates the datasets used and shows the experimental results and their analysis. Finally, Section V provides the conclusion and future directions.

II. Sparse Representation Classifiers

Given an HSI $\mathbf{H} \in \mathbb{R}^{P\times Q\times B}$, where $P \times Q$ is the size of the image and B is the number of bands, the SRC strategy involves finding a linear combination of atoms (labeled pixels) for an unknown pixel $y_j \in \mathbb{R}^B$, such that

$$\mathrm{y_j} = \mathbf{D}\alpha \tag{1}$$

where $\mathbf{D} = [\mathbf{D}_1, \mathbf{D}_2, \ldots, \mathbf{D}_C]$ is a $B \times N$ dictionary, where $\mathbf{D}_i$ is a subset of the dictionary containing $N_i = |D_i|$ atoms of the ith class; hence, $N = \sum_{i=1}^{C} N_i$, and α is a sparse vector whose nonzero entries represent the weights of particular atoms.

The KSRC performs classification by casting the original SRC problem onto higher dimensional feature space [19]. The kernel function $\Phi : \mathcal{X} \to \mathcal{F}$ projects a pixel of HSI from its original feature space to kernel space. Thus, the unknown pixel y_j and the dictionary D can be represented in the kernel space as $\Phi(y_j)$ and $V = [\Phi(x_1), \Phi(x_2), \ldots, \Phi(x_N)]$, respectively, where $\Phi(x_l)$, $l = 1, 2, \ldots, N$ represents the atom (training sample) in higher dimensional kernel space, and let α' be the new sparse vector. Therefore, the sparse representation of y_j in kernel space becomes

$$\Phi(y_j) = V\alpha'. \tag{2}$$

KSRC follows a two-step procedure. The first step involves finding α' by solving the following optimization problem using greedy kernel orthogonal matching pursuit (KOMP)

$$\hat{\alpha}' = \arg\min \|\Phi(\mathrm{y_j}) - \mathrm{V}\alpha'\|_{\mathrm{F}} \quad \text{s.t. } \|\alpha'\|_0 < K \tag{3}$$

where K represents the sparsity level and $\|.\|_F$ is the Frobenius norm. Then, in the second step, the residue between y_j and its approximation for each class is calculated. Finally, the class label that provides the minimum residue for y_j is assigned to it, i.e.,

$$f_i(y_j) = \|\Phi(\mathrm{y_j}) - V_i\hat{\alpha}_i'\|_F, \quad \text{for } i = 1, 2, \ldots, C \tag{4}$$

$$\text{class}(y_j) = \arg\min_{i=1,2,\ldots,C} \{f_i(y_j)\}. \tag{5}$$

In order to consider spatial contextual information, an extended version of KSRC known as kernel joint SRC (KJSRC) has proposed in [20]. It assumes that neighboring pixels of an unknown pixel can be represented by linear combination of the same atoms. Hence, their sparse vectors share the same sparsity pattern but with different coefficients. For details about KJSRC, the reader may refer to [20].

III. Proposed Method

In this section, we propose a query function consists of uncertainty and diversity criteria designed by exploiting the KSRC. The query function of the proposed technique selects the h most informative samples at each iteration of AL by applying uncertainty criterion followed by diversity criterion. The details of these criteria are presented in Sections III-A and III-B.

A. Uncertainty Criterion

In AL, the uncertainty criterion plays a major role in selecting informative samples. As already mentioned, in the literature, most of the AL techniques exploited SVM classifier to define uncertainty criteria. SRC, a robust and important classifier for HSI classification, has been used for AL to some extent. However, they have not achieved convincing results as compared with the existing SVM-based AL models. The BT criterion defined in [17] fails to find out appropriate uncertain samples, since the HSI samples that belong to different classes are often highly overlapped. In order to improve the separation between overlapping classes, the KSRC that projects the samples into a higher dimensional space is a better choice. Accordingly, we introduce an uncertainty criterion, called kernel BT (KBT), that is defined by exploiting the KSRC. Given an unlabeled sample, the reconstruction error for each class is first computed using the dictionary of KSRC. Then, the classification confidence, i.e., the difference between lowest and second-lowest reconstruction errors, is used to define the uncertainty level of the sample. The classification confidence $E_{\text{kbt}}(y_j)$ for a sample y_j in the unlabeled pool U is computed as follows:

$$l_1 = \arg\min_{i=1,2,\ldots,C} \{f_i(y_j)\}, \quad l_2 = \arg\min_{i=1,2,\ldots,C,\, i\neq l_1} \{f_i(y_j)\}$$
$$E_{\text{kbt}}(y_j) = f_{l_2}(y_j) - f_{l_1}(y_j). \tag{6}$$

Lower values of $E_{\text{kbt}}(y_j)$ imply that the sample y_j lies in the boundary region between two classes and is more uncertain. Higher values of $E_{\text{kbt}}(y_j)$ imply that the sample y_j is less uncertain. The proposed technique selects the batch of uncertain samples from U, which have the lowest classification confidence as computed in (6).

B. Diversity Criterion

In batch mode AL, high redundancy in the uncertainty samples affects the selection of informative samples. In order to select informative samples (i.e., uncertain as well as nonredundant) at each iteration of batch mode AL, most of the literature methods first apply an uncertainty criterion to select m uncertain samples (may have redundancy) from U. Then, a diversity criterion is used to select h informative samples from the m samples for labeling. In the literature, a large number of diversity criteria, such as closest SV, clustering-based diversity, angle-based diversity, and so on, are exist [6]. We propose a new diversity criterion called dictionary correlation-based diversity (DCBD) by exploiting the dictionary of KSRC. The DCBD selects samples that have minimum redundancy with the dictionary by measuring the correlation between uncertain samples and the atoms in the dictionary.

Our AL technique first selects m samples $U_m = [y_1, y_2, \ldots, y_m]$ from U using the proposed uncertainty criterion. Then, a first-order sequential search is performed that selects one sample at a time from U_m, having minimum redundancy with the atoms in the dictionary of KSRC. It selects $h (h < m)$ samples from the m uncertain samples. Let $V = [\Phi(x_1), \Phi(x_2), \ldots, \Phi(x_r)]$ be the initial dictionary of KSRC. For each sample, $y_i \in U_m$ the correlation values between $\Phi(y_i)$ and all the atoms in V are computed. Let $\langle V, \Phi(y_i)\rangle$ be the vector that stores m correlation values computed between $\Phi(y_i)$ and r atoms in V. We define a function $R(y_i) = \max\{< V, \Phi(y_i) >\}$ that takes the maximum correlation value from the vector $\langle V, \Phi(y_i)\rangle$ to measure the redundancy of $\Phi(y_i)$ in the dictionary V. High values of $R(y_i)$ imply that the sample $\Phi(y_i)$ will be a redundant atom in V. Thus, in the proposed technique, the sample $y_k \in [y_1, y_2, \ldots, y_m]$ that has the least redundancy with the atoms in V is selected as follows:

$$y_k = \underset{y_1, y_2, \ldots, y_m}{\arg\min} \{R(y_1), R(y_2), \ldots, R(y_m)\}. \tag{7}$$

After assigning an appropriate label for y_k, the dictionary of KSRC and U_m are updated as $V = V \cup \{\Phi(y_k)\}$ and $U_m = U_m - \{y_k\}$. The above sequential search is repeated for h times to add h atoms into the dictionary of KSRC with minimum redundancy.

The whole process of the selection of h uncertain and diverse samples for updating the dictionary is repeated until the stopping criterion is met. Finally, the obtained dictionary is used by KJSRC for classification. Algorithm 1 summarizes the proposed AL framework with KBT uncertainty criterion and DCBD diverse criterion. The main theoretical motivation of the proposed technique is that uncertainty and diversity criteria are designed in the kernel sparse representation and not in spaces different from those in which KSRC works. In this way, it is possible to optimize the selection of samples that are the most uncertain and the most diverse in the considered kernel sparse representation space used by the classifier.

Algorithm 1 Proposed AL Strategy Using KBT Uncertainty Criterion and DCBD Diversity Criterion

Input: Hyperspectral Image **H**; sparsity level K; Dictionary **D**; unlabelled pool U

Let $V = [\Phi(x_1), \Phi(x_1), \ldots, \Phi(x_r)]$ be the initial dictionary of KSRC
repeat:
For each $y_i \in U$, compute classification confidence $E_{kbt}(y_i)$ using (6)
Select the m uncertain samples U_m from U that have the lowest classification confidence.
$p = m$
/* To select h informative samples */
for i=1 to h **do**
  for j=1 to p **do**
    Compute Correlation Vector $< V, \Phi(y_j) >$
    $R(y_j) = max\{< V, \Phi(y_j) >\}$
  end for
  $y_k = \underset{y_1, y_2, \ldots, y_m}{\arg\min} \{R(y_1), R(y_2), \ldots, R(y_m)\}$
  Assign appropriate class label to y_k with the help of supervisor.
  Update:
  $V = V \cup \Phi(y_k);\ U_m = U_m - \{y_k\};\ U = U - \{y_k\}$
  $p = p - 1$
end for
until stopping criterion is met.

IV. Experimental Results

A. Datasets

We assessed the effectiveness of the proposed technique on three benchmark datasets.

1) Indian Pines: This dataset was acquired by the AVIRIS sensor on the Indian Pines, IN, USA. The image scene contains 200 bands and 145 × 145 pixels. There are 16 ground-truth classes present in it.

2) Kennedy Space Center: This dataset was taken by the AVIRIS sensor on the Kennedy Space Center (KSC), FL, USA. The image scene contains 176 bands and 512 × 614 pixels. There are 13 ground-truth classes present in it.

3) University of Houston: This dataset was taken by the CASI sensor. It shows the University of Houston, TX, USA, and the surrounding urban areas. The image scene contains 144 bands and 349 × 1905 pixels. There are 15 ground-truth classes present in it.

B. Design of Experiments

In order to show the robustness of our proposed technique, two sets of experiments are conducted, one to show the potentiality of the proposed KBT criterion for selecting uncertain samples and the other to show the potentiality of the proposed DCBD criterion for selecting informative samples. In the first experiment, to assess the effectiveness of the proposed uncertainty criterion, the KBT criterion is compared with several popular state-of-the-art uncertainty criteria, such as MS [4], MCLU [6], EQB [5], and sparse representation-based BT (SRC-BT) [17]. In the second experiment, to assess the effectiveness of the proposed diversity criterion, the KBT-DCBD technique is compared with three state-of-the-art AL techniques, i.e., MS with closest SV (MS-cSV) [5], MCLU with kernel k-means (MCLU-KK) [6], and MCLU-KK with edge gradient (MCLU-KK-EG) [12]. The classification results of the proposed technique not change significantly by varying the sparsity level K of KSRC and the γ of RBF kernel within the range $[3, 9]$ and $[2^5, 2^8]$, respectively. For the present experiment, the value of K is fixed to 9 for the University of Houston dataset and to 3 for the

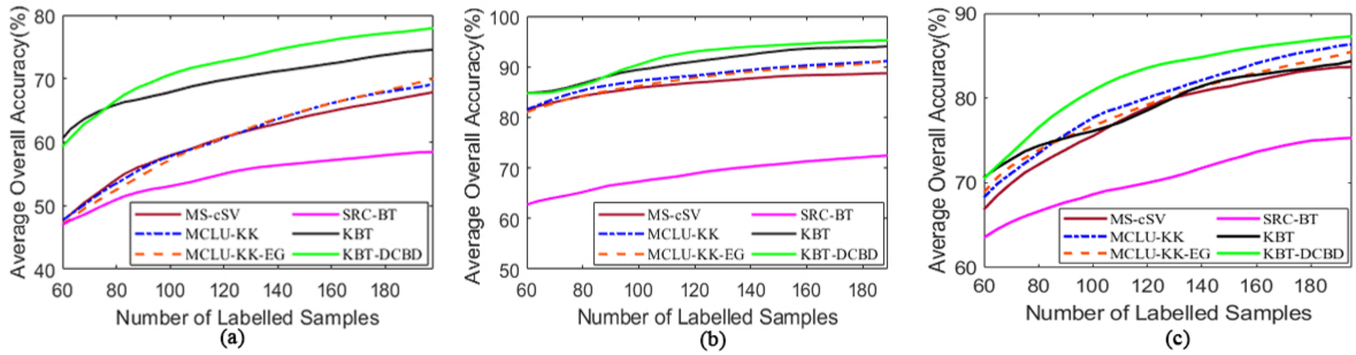

Fig. 1. ($\overline{\text{OA}}$) versus the number of labeled samples of different AL methods for (a) Indian Pines, (b) KSC, and (c) University of Houston datasets.

other datasets. The value of γ is fixed to 2^7, and the sparsity level of KJSRC is set to 30 for all the datasets.

At the beginning of AL, three samples from each class are randomly selected to form the initial training set (dictionary). Subsequently, in each iteration of AL, $h = 5$ samples are selected and assigned to their appropriate labels for updating the training set (dictionary). The MS-cSV, MCLU-KK, MCLU-KK-EG, and the KBT-DCBD techniques first select m uncertain samples. Then, h informative samples from the selected m samples are chosen for labeling. In the experiments, the value of m is set to $3h$. The AL techniques are executed for 30 iterations, resulting in 198, 189, and 195 samples in the training set (dictionary) for the Indian Pines, KSC, and University of Houston datasets, respectively. The rest of the samples in U (i.e., 10 051, 4838, and 14 839 samples for the Indian Pines, KSC, and University of Houston datasets) are used for testing. Experiments are performed in five runs to reduce random effects on the results.

All experiments are performed in a 64-bit MATLAB R2018a environment on a desktop with a 64-bit operating system, Intel[1] Core[2] i5 8400 CPU, 2.80-GHz processing power, and 16 GB of RAM.

C. Results

Table I shows the results of the first experiment obtained at different iterations of AL for Indian Pines, KSC, and University of Houston, respectively. It shows the average overall accuracy ($\overline{\text{OA}}$) and the standard deviation (sd) provided by all the considered techniques with different numbers of labeled (training) samples N. From the table, one can see that for all the datasets, the proposed uncertainty criterion (i.e., KBT) outperforms many popular state-of-the-art uncertainty criteria that exist in the AL literature. For the Indian Pines dataset, the proposed criterion provides at least 6% higher average overall accuracy ($\overline{\text{OA}}$) than the best literature criterion (i.e., MCLU). Similarly, for the KSC dataset, the KBT provides nearly 4% higher $\overline{\text{OA}}$ than the best literature criterion. Also, for the University of Houston dataset, the proposed criterion provides either better or similar results as provided by the best literature criterion (i.e., MCLU). Thus, the results show

TABLE I

AVERAGE OVERALL ACCURACY ($\overline{\text{OA}}$) AND STANDARD DEVIATION (SD) OBTAINED BY THE MS [4], MCLU [5], EQB [5], SRC-BT [17], AND KBT FOR THE INDIAN PINES, KSC, AND UNIVERSITY OF HOUSTON DATASETS

Method	$N=88$		$N=128$		$N=163$		$N=198$	
Indian Pines								
	$\overline{OA}$	sd	$\overline{OA}$	sd	$\overline{OA}$	sd	$\overline{OA}$	sd
MS	56.41	± 3.85	61.80	± 2.53	65.89	± 2.10	68.37	± 1.79
MCLU	57.47	± 2.36	60.97	± 1.48	65.50	±1.25	68.66	± 1.19
EQB	49.46	± 5.67	55.83	± 4.81	59.57	± 3.62	62.16	± 2.21
SRC-BT	52.78	± 1.51	55.79	± 1.91	57.33	± 1.49	58.45	± 1.40
KBT	66.76	± 4.33	70.65	±1.84	72.54	± 2.04	74.56	±1.75
Kennedy Space Center								
Method	$N=79$		$N=119$		$N=154$		$N=189$	
	$\overline{OA}$	sd	$\overline{OA}$	sd	$\overline{OA}$	sd	$\overline{OA}$	sd
MS	84.84	± 1.64	86.89	± 0.86	87.94	± 0.86	89.10	± 0.70
MCLU	84.71	± 2.09	87.69	± 0.93	89.51	± 0.71	90.43	± 0.28
EQB	82.81	± 1.62	85.74	±1.63	87.54	± 0.85	88.65	± 1.18
SRC-BT	65.29	± 2.82	68.12	±4.01	69.65	± 4.49	72.85	± 2.66
KBT	87.77	± 2.33	90.43	± 2.96	93.49	± 1.53	94.05	± 1.22
University of Houston								
Method	$N=80$		$N=135$		$N=160$		$N=195$	
	$\overline{OA}$	sd	$\overline{OA}$	sd	$\overline{OA}$	sd	$\overline{OA}$	sd
MS	70.32	± 2.84	77.32	± 1.72	82.93	± 0.90	83.47	± 0.72
MCLU	72.63	± 1.46	78.99	± 1.49	81.98	± 1.01	84.58	± 0.65
EQB	68.25	± 1.92	73.73	± 1.24	76.58	± 2.30	78.97	± 1.98
SRC-BT	66.86	± 1.23	69.86	± 0.94	73.57	± 1.49	75.31	± 1.65
KBT	74.34	± 1.22	80.91	± 2.81	82.77	± 2.35	84.38	± 2.39

the potentiality of the proposed KBT criterion to select better uncertain samples.

The results of the second experiment for Indian Pines and KSC are shown in Fig. 1. From these figures, one can see that our KBT provides significantly better results than many other state-of-the-art methods. It is also seen that the KBT-DCBD method further improves the results. Note that only at the beginning, the KBT-DCBD failed to provide better results than KBT. This is because at the initial iterations of AL, the decision function of the classifier may be passed through completely wrong regions in the feature space. So, the diverse samples selected may not be useful to provide additional information to the classifier. For the Indian Pines dataset, our KBT-DCBD technique provides an $\overline{\text{OA}}$ of 77.94%

[1]Registered trademark.
[2]Trademarked.

TABLE II

COMPUTATIONAL TIME IN SECONDS TAKEN BY THE MS-cSV, MCLU-KK, MCLU-KK-EG, SRC-BT, KBT, AND KBT-DCBD FOR DIFFERENT DATASETS

Dataset	MS-cSV	MCLU-KK	MCLU-KK-EG	SRC-BT	KBT	KBT-DCBD
IP	395.75	131.08	138.98	398.13	174.99	175.92
KSC	190.03	49.64	51.08	386.12	73.93	74.33
UH	491.17	134.98	135.84	521.56	242.87	243.92

by labeling only 198 samples, whereas with the same number of labeled samples, the best literature method (i.e., the MCLU-KK-EG) provides only 69.95%. For the KSC dataset, our technique provides an $\overline{\text{OA}}$ of 95.22% by labeling only 189 samples, whereas the best literature method (i.e., the MCLU-KK) obtains 91.14%. Similarly, for the University of Houston (UH) dataset, the proposed KBT-DCBD technique provides an $\overline{\text{OA}}$ of 87.29%, whereas the best literature method MCLU-KK provides 86.35% of $\overline{\text{OA}}$. These improvements are made possible from the sparse representation models in kernel space, which are exploited by the proposed technique for designing better uncertainty and diversity criteria.

The computational time values taken by the different techniques are reported in Table II. From this table, one can see that the computational time of the proposed technique is comparable to the literature methods.

V. CONCLUSION

The goal of this research is to develop a novel robust query function for improving the effectiveness of AL in identifying appropriate informative samples to label for the training of a KSRC. The proposed uncertainty and diversity criteria are designed in the kernel sparse representation and not in spaces different from those in which KSRC works. So, it is possible to optimize the selection of samples that are the most uncertain and the most diverse in the considered kernel sparse representation space used by the classifier with the results to obtain a very good trade-off between labeling cost and classification accuracy. The robustness of the proposed technique is validated by using three real benchmark HSI datasets. For all the datasets, the proposed method improves the classification accuracy with a reduced number of labeled training data. This is very important and goes in the direction of improving the real-life deployment of an automatic classifier as results in a reduction of the cost for labeling training data with respect to other literature methods. Moreover, the proposed method does the following: 1) is based on a simple design; 2) is fast; 3) does not have critical requirements on memory; and 4) exploits a sparse classifier robust to high-dimensional feature spaces.

REFERENCES

[1] Y. Zhang, K. Liu, Y. Dong, K. Wu, and X. Hu, "Semisupervised classification based on SLIC segmentation for hyperspectral image," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 8, pp. 1440–1444, Aug. 2019.

[2] S. Patra and L. Bruzzone, "A batch-mode active learning technique based on multiple uncertainty for SVM classifier," *IEEE Geosci. Remote Sens. Lett.*, vol. 9, no. 3, pp. 497–501, May 2012.

[3] X. Cao, J. Yao, Z. Xu, and D. Meng, "Hyperspectral image classification with convolutional neural network and active learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 7, pp. 4604–4616, Jul. 2020.

[4] P. Mitra, B. U. Shankar, and S. K. Pal, "Segmentation of multispectral remote sensing images using active support vector machines," *Pattern Recognit. Lett.*, vol. 25, no. 9, pp. 1067–1074, 2004.

[5] D. Tuia, F. Ratle, F. Pacifici, M. F. Kanevski, and W. J. Emery, "Active learning methods for remote sensing image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 7, pp. 2218–2232, Jul. 2009.

[6] B. Demir, C. Persello, and L. Bruzzone, "Batch-mode active-learning methods for the interactive classification of remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 3, pp. 1014–1031, Mar. 2011.

[7] S. Patra and L. Bruzzone, "A fast cluster-assumption based active-learning technique for classification of remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 5, pp. 1617–1626, May 2011.

[8] E. Pasolli, F. Melgani, D. Tuia, F. Pacifici, and W. J. Emery, "SVM active learning approach for image classification using spatial information," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 4, pp. 2217–2233, Apr. 2014.

[9] K. Brinker, "Incorporating diversity in active learning with support vector machines," in *Proc. 20th Int. Conf. Mach. Learn.*, 2003, pp. 59–66.

[10] S. Patra and L. Bruzzone, "A novel SOM-SVM-based active learning technique for remote sensing image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 11, pp. 6899–6910, Nov. 2014.

[11] S. Patra, K. Bhardwaj, and L. Bruzzone, "A spectral–spatial multicriteria active learning technique for hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 12, pp. 5213–5227, Dec. 2017.

[12] A. Samat, J. Li, C. Lin, S. Liu, and E. Li, "Edge gradient-based active learning for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 9, pp. 1588–1592, Sep. 2020.

[13] C. Mu, J. Liu, Y. Liu, and Y. Liu, "Hyperspectral image classification based on active learning and spectral–spatial feature fusion using spatial coordinates," *IEEE Access*, vol. 8, pp. 6768–6781, 2020.

[14] M. Xu, Q. Zhao, and S. Jia, "Multiview spatial–spectral active learning for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5512415, doi: 10.1109/TGRS.2021.3095292.

[15] D. Li, Q. Wang, and F. Kong, "Adaptive kernel sparse representation based on multiple feature learning for hyperspectral image classification," *Neurocomputing*, vol. 400, pp. 97–112, Aug. 2020.

[16] H. Yu, X. Shang, X. Zhang, L. Gao, M. Song, and J. Hu, "Hyperspectral image classification based on adjacent constraint representation," *IEEE Geosci. Remote Sens. Lett.*, vol. 18, no. 4, pp. 707–711, Apr. 2021.

[17] L.-Z. Huo, L.-J. Zhao, and P. Tang, "Hyperspectral image classification with sparse representation classifier and active learning," in *Proc. 8th Workshop Hyperspectral Image Signal Process., Evol. Remote Sens. (WHISPERS)*, Aug. 2016, pp. 1–5.

[18] G. Wang, J.-N. Hwang, C. Rose, and F. Wallace, "Uncertainty-based active learning via sparse modeling for image classification," *IEEE Trans. Image Process.*, vol. 28, no. 1, pp. 316–329, Jan. 2019.

[19] Y. Chen, N. M. Nasrabadi, and T. D. Tran, "Hyperspectral image classification via kernel sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 1, pp. 217–231, Jan. 2013.

[20] E. Zhang, X. Zhang, L. Jiao, H. Liu, S. Wang, and B. Hou, "Weighted multifeature hyperspectral image classification via kernel joint sparse representation," *Neurocomputing*, vol. 178, pp. 71–86, Feb. 2016.