Entropy query by bagging-based active learning approach in the extreme learning machine framework for hyperspectral image classification

Monoj K. Pradhan[1,*], **Sonajharia Minz**[1] **and Vimal K. Shrivastava**[2]

[1]School of Computer and Systems Sciences, Jawaharlal Nehru University, New Delhi 110 067, India
[2]School of Electronics Engineering, Kalinga Institute of Industrial Technology, Bhubaneswar 751 024, India

Active learning (AL) technique is the classification of remote sensing images, where collecting efficient training data is costly in terms of labour and the time taken. The prime objective of AL technique is to obtain high classification accuracy with the training sample as compact as possible. Most studies on the classification of remote sensing images using AL, focused only on accuracy, with hardly any study on computation time. Keeping reduction of computation time as the objective, here we present, an entropy query by bagging (EQB)-based AL approach in the extreme learning machine (ELM) framework for remote sensing image classification. The performance of this approach is compared with the widely used support vector machine (SVM) AL framework in combination with different query strategies. To verify the efficacy of the study, the approaches were tested on two hyperspectral remote-sensing images, i.e. Kennedy Space Centre (KSC) and Botswana (BOT). The proposed system depicts competitive classification performance while significantly reducing computation time.

Keywords: Active learning, computation time, extreme learning machine, entropy query by bagging.

As a significant breakthrough in the area of remote sensing images in the past, the hyperspectral image (HSI) is capable of acquiring hundreds of narrow spectral bands varying from visible to infrared. These spectral bands carry information that has greater ability to analyse land-cover scenes on the earth's surface, which is difficult in multispectral images[1]. These features of HSI provide scope for studying different applications such as urban planning[2], forest monitoring[3], agriculture[4] and land cover[5]. HSI classification is one of the basic features in many remote sensing applications and has attracted the attention of several researchers recently[6–10]. Many supervised learning algorithms have been explored to solve the classification problems. The performance of supervised learning is influenced by the quality and quantity of the training set. However, obtaining an efficient training set is costly in terms of labour and time in case of remote sensing images[11]. To address such issues, the active learning (AL) technique is extensively used where only uncertainty and representative training samples are selected to train the model[11–18].

The AL algorithm is a strategy that collects most uncertainty samples from the unlabelled dataset by applying a query strategy[19]. The true class labels are assigned to these uncertainty samples to retrain the model with an updated labelled dataset, thus improving the classification performance. Hence, there is no need to label non-uncertainty samples, which effectively decreases the time and cost[11]. However, determining the uncertainty samples is a challenging task. Hence, the main focus of the AL technique is query strategy, which finds the samples with most uncertainty from the unlabelled dataset to increase the learning ability of the classifier. The various query strategies are presented in the literature[19,20]. They are grouped mainly into three families, viz. (i) large margin-based strategy, (ii) posterior-based strategy and (iii) committee-based strategy. The first family is mainly based on the support vector machine (SVM) classifier[19]. Various heuristics, in this family, include, margin sampling (MS)[21], significance space construction (SSS)[22] and multiclass level uncertainty (MCLU)[23]. The second family relies on the concept of posterior probability, and includes classifiers such as maximum likelihood classifier[13], multinomial logistic classifier[24,25] and probabilistic SVM classifier[19,26] with AL heuristic such as Kullback–Leibler (KL)–Max[13], mutual information (MI)[27] and breaking ties (BT)[28]. The third family can adopt any type of classifiers. The key feature of the third family is to choose samples with the most uncertainty on the basis of heuristics that includes active democratic co-learning[29], query by bagging[30], active DECORATE[31], entropy based AL[20] and multiview disagreement-based AL[11,32].

According to the literature, enough classification accuracy has been gained for HSI classification utilizing the AL approach, but the computing time is very high[33,34]. However, it is observed that the extreme learning

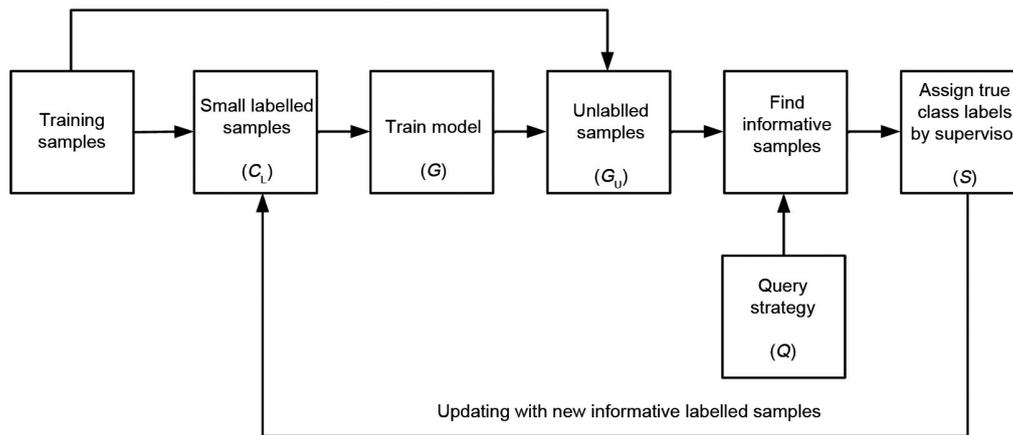*For correspondence. (e-mail: monojpradhan76@gmail.com)

Figure 1. Block diagram of the general active learning process.

machine (ELM) gives equally good classification accuracy in comparison to traditional classifiers, while significantly reducing the computation time[35–37]. ELM has been explored for HSI classification[38–41], but there are only a few attempts to classify HSI using ELM with the AL technique[16,42]. This study discusses entropy query by bagging (EQB)-based AL approach in ELM framework for the classification of HSI. Moreover, the performance of this approach has been compared with the widely used support vector machine (SVM) AL framework, in combination with different query strategies such as random sampling (RS) and multiview (MV). In this way, we have compared six AL models: ELM–RS, ELM–MV, ELM–EQB, SVM–RS, SVM–MV and SVM–EQB. Thus, the objectives of this study are: (i) integrating EQB query strategy into the AL technique with ELM classifier; (ii) comparative performance analysis of six AL models made up of crisscross combination of two classifiers (SVM, ELM) and three query strategies (RS, MV, EQB) and (iii) obtaining adequate classification accuracy using the proposed model (ELM–EQB) while significantly reducing the computation time.

Methodology

AL is a strategy to iteratively collect uncertainty samples from the unlabelled dataset through a trained model. The uncertainty samples are also referred to as the most informative samples. The AL model is a quintuple (G, Q, S, C_L and C_U)[43]. Samples from the dataset are unbiasedly grouped as training and testing (C_T) sets. The training set is grouped into a small set (C_L) having labels called initial training set and the remaining as unlabelled set (C_U). G denotes the classifier which is first trained with the initial C_L. While Q is a query function which is utilized to collect samples with the most uncertainty from C_U. S acts as a supervisor and plays a vital role in assigning the correct class labels to the collected uncertainty samples.

Collecting uncertainty samples, updating of C_L by appending these uncertainty samples and retraining of G with the updated G_L are performed at every iteration. This process is repeated till the predefined condition (number of iterations) is met. In this study, ELM is used as a classifier (G) and EQB is used as a query strategy (Q). Figure 1 is a block diagram of general AL process.

Extreme learning machine

The ELM, first proposed by Huang et al.[35], is a single hidden layer feed forward neural network (SLFN). The pertinent parameters such as weights and biases are randomly assigned from the input to the hidden layer and not tuned in the entire process. The weights and biases are updated only in the hidden to the output layer. In this way, ELM achieves excellent generalization performance, fast computation and straight-forward solution[35–37]. Figure 2 shows the architecture of ELM. It comprises N, L and M number of neurons in the input layer, hidden layer and output layer respectively. Let us consider C_L training samples, each denoted as (x_j, y_j) where x_j and y_j are the input and output vector of the jth sample. The ELM model with an activation function $G(w_i, b_i, x_j)$ can be expressed as

$$f_L(x_j) = \sum_{i=1}^{L} \beta_i G(w_i, b_i, x_j); \ j = 1, 2, ..., C_L$$

$$\text{and } i = 1, 2, .., L, \tag{1}$$

where w_i and b_i are the ith weight vector and bias from the input layer to the hidden layer respectively, and β_i is the weight vector connecting the ith hidden layer to the output layer. Here, $G(w_i, b_i, x_j)$ denotes the activation function (sigmoid, radial basis, sine, etc.) which is used to determine the output of the ith hidden node for the jth training sample. Each neuron of the input layer is connected to all neurons of the hidden layer.

Here, the focus is to minimize the cost function (E) of ELM using eq. (2) as given below

$$E = \sum_{j=1}^{N} \left(\sum_{i=1}^{L} \beta_i G(w_i, b_i, x_j) - y_j \right)^2. \qquad (2)$$

Query strategy

The strategy to determine the samples with most uncertainty is known as the query strategy. Here, we have used the EQB strategy, and compared it with RS and MV as explained in the following sections.

Random sampling-based active learning: RS-based AL (RS–AL) randomly picks P number of samples (C_P) from the unlabelled dataset (C_U) without considering any criteria. These samples are removed from C_U and appended to C_L after assigning true labels by a supervisor (S). Further, the classifier (G) is retrained on the updated C_L. The algorithm 1 in Box 1 presents the framework of RS–AL.

Multiview-based active learning: The classification of HSI using the MV query strategy was first introduced by Di and Crawford[11]. Simplicity and flexibility are the advantages of this strategy compared to different existing query strategies. The spectral features in the MV scenario are grouped into several disjoint subsets called views $(X_1 \times X_2 \times \cdots X_{K_V})$; where K_V denotes the number of views and X the sample space in HSI. Here we assume that each view is sufficient to train the model. The key idea of this strategy is based on disagreement between the trained models. The samples with most uncertainty were collected based on maximum disagreement between the

trained models for different views. Adaptive maximum disagreement (AMD) technique is used to compute the disagreement between the trained models[11] and is given in eq. (3) below.

$$D_{\max} = \max_{x_i \in C_U} D(x_i, f_V^1, f_V^2, \ldots f_V^{K_V}), \qquad (3)$$

where $D_{\max}$ is the maximum disagreement value, x_i is a sample from C_U, f_V^j is the jth view and $D(\cdot)$ is the disagreement value on each x_i between different views and can be defined as follows:

$$D(x_i, f_V^1, f_V^2, \ldots, f_V^{K_V}) = \text{count} \, | \, f_V^j \, | \quad \text{for} \quad j = 1, 2, \ldots, K_V, \qquad (4)$$

where count $|\cdot|$ is used to count the number of unique elements in the set. Algorithm 2 in (Box 2) presents the framework of MV–AL.

Entropy query by bagging-based active learning: Seung et al.[44] proposed the query by committee (QBC) strategy based on maximum disagreement. Freund et al.[45] used this strategy to collect the uncertainty sample from a random stream of inputs, while Abe and Mamitsuka[30] used the QBC model in binary classification. Tuia et al.[20] extended this concept on the multiclass classification using entropy heuristic known as entropy query by bagging. The focus of EQB is to build k training sets on bootstrap samples (C'_{L_k})[46]. A bootstrap is formed with replacement of the original samples in every iteration. Each set of bootstrap (C'_{L_k}) contains a subset of C_L. C'_{L_k} is formed by picking defined percentage (pct) of samples randomly from C_L. k number of bootstrap sets form a
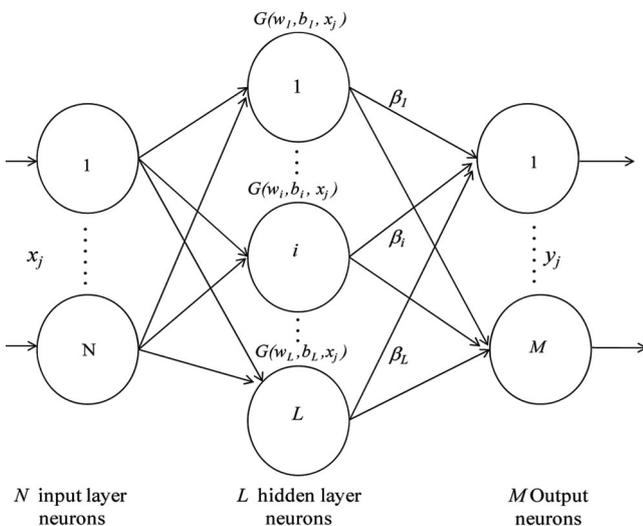

Figure 2. Architecture of the extreme learning machine (ELM) followed in this study.

Box 1.

Algorithm 1: Framework of RS–AL

Initialization steps:
 1: Randomly divide the dataset into initial training (C_L), unlabelled (C_U) and test (C_T) sets.
 2: Initialize the number of samples (P) to be chosen at each iteration.
 3: Initialize the number of iterations (iter) to be performed.
 4: Train the classifier (G) on C_L.

Repeat

 5: Test the trained classifier (G) on C_T and obtain the classification accuracy.
 6: Randomly pick P samples (C_P) from C_U.
 7: Extract C_P from C_U and remove it from C_U.
 8: Label C_P by supervisor (S).
 9: Append C_P to C_L.
 10: Retrain G with updated C_L.
 11: iter = iter + 1.

Until stopping criteria (number of iterations) is satisfied.

committee of k models. These bootstrap sets train the corresponding models, and the trained models are reused to determine the class label of unlabelled samples (C_U). Therefore, k possible labels for each x_i were predicted from C_U. The selected uncertainty samples are the ones for which the predictions are most evenly split as shown in eq. (5) below for binary classification.

$$\hat{x} = \arg \min_{x_i \in C_U} \| \{G \leq k | f_G(x_i) = 1\} - \{G \leq k | f_G(x_i) = 0\} \|, \tag{5}$$

where G is one among the committee of models. If the models agree to a certain prediction, then eq. (5) is maximized. On the contrary, the uncertainty samples will produce small values in the equation. Tuia et al.[20] extended eq (5) to a multiclass problem using maximum entropy distribution heuristic of the models. The entropy distribution $H(x_i)$ is calculated from the k labels of x_i as follows:

$$H(x_i) = \sum_k -p_{i,k} \log_{10}(p_{i,k}), \tag{6}$$

where $p_{i,k}$ denotes the probability of having class k for $x_i \in C_U$. $H(x_i)$ is calculated for each $x_i \in C_U$. The maximum entropy is calculated as follows:

$$\hat{x} = \max_{x_i \in C_U} H(x_i). \tag{7}$$

The samples with maximum disagreement value between models resulted in maximum entropy. C_P was removed from C_U and added to C_L. Algorithm 3 in Box 3 describes the EQB-based AL. Figure 3 shows the flow diagram of EQB-based AL approach.

Experimental design

Data description

In order to show the effectiveness of the proposed technique, two HSI datasets, viz. Kennedy Space Centre (KSC) and Botswana (BOT) were used in the experiment[47].

Box 2.

Algorithm 2: Framework of MV–AL

Initialization steps:
1: Randomly divide the dataset into initial training set (C_L), unlabelled set (C_U) and test set (C_T).
2: Initialize the number of samples (P) to be selected at each iteration.
3: Initialize the number of iterations (iter) to be performed.
4: Generate K_V views ($X_1, X_2, ..., X_{K_V}$) from the available spectral features (X) of HSI.
5: Construct the classification models ($G_1, G_2, ..., G_{K_V}$) corresponding to the views ($X_1, X_2, ..., X_{K_V}$).

Repeat

6: Test the trained classifier on C_T and obtain the classification accuracy.
7: Predict the class label of each $x_i \in C_U$ corresponding to the views ($X_1, X_2, ..., X_{K_V}$).
8: Determine the disagreement value (D) of $x_i \in C_U$ using eq. (4).
9: Arrange samples in C_U in descending order on the basis of disagreement values.
10: Extract C_P having maximum disagreement value using eq. (3) and discard these samples from C_U.
11: Assign true class label to the retrieved C_P by supervisor (S) and update C_L by appending C_P.
12: Retrain the models ($G_1, G_2, ..., G_{K_V}$) with updated C_L.
13: iter = iter + 1.

Until stopping criteria (number of iterations) is satisfied.

Box 3.

Algorithm 3: Framework of EQB–AL

Initialization steps:
1: Randomly divide the dataset into the initial training set (C_L), unlabelled set (C_U) and test set (C_T).
2: Initialize the number of samples (P) to be selected at each iteration.
3: Initialize the number of iterations (iter) to be performed.
4: Initialize a set of bootstrap samples (k).
5: Initialize percentage (pct) of samples to be drawn from C_L for bootstrap samples.
6: Train the model (G) with the current set of training samples (C_L).

Repeat

7: Compute the classification accuracy on C_T using G.

for $G = 1$ to k do

8: Obtain subset (C'_{L_G}) by applying pct of C_L.
9: Train the Gth model using C'_{L_G}.

End of for

10: Predict the class membership of each $x_i \in C_U$ by each Gth model.
11: Compute the entropy for every $x_i \in C_U$ using eq. (6).
12: Extract samples C_P form C_U with maximum entropy using eq. (7) and assign to further label by supervisor (S).
13: Update C_L by appending the extracted C_P.
14: Retrain G with the updated G_L.
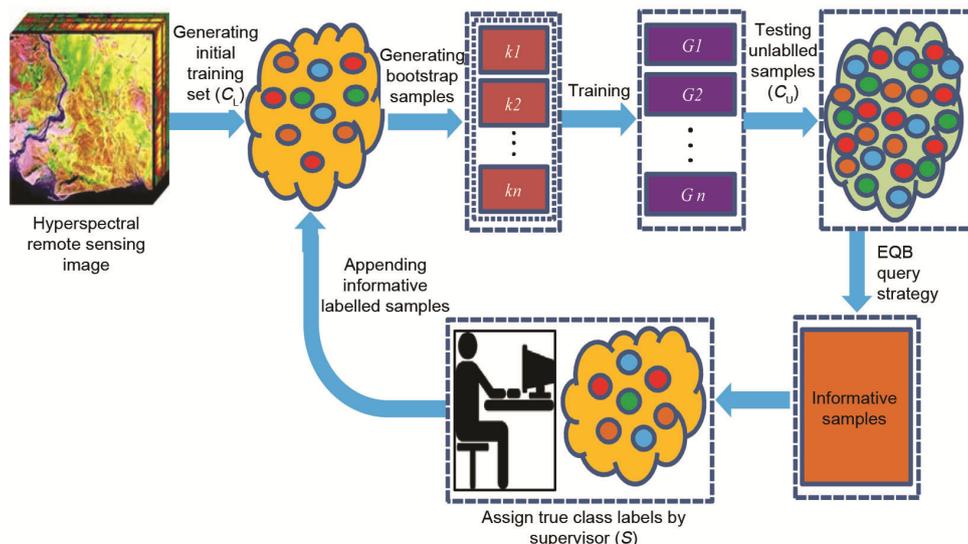
Until stopping criteria is satisfied.

Figure 3. Flow diagram of entropy query by bagging (EQB) based active learning approach.

These datasets have already been pre-processed by the Kennedy Space Center and the UT Center for Space Research, USA respectively[47]. Both datasets have several classes with similarity of vegetation signatures[11]. Therefore, it is a challenging task to get high classification performance on these datasets, which motivated us to choose them. The KSC dataset[33], shown in Figure 4 a, was acquired from KSC, Florida, USA on 23 March 1996. It comprises 224 spectral bands and was reduced to 176 spectral bands after removal of noisy and water bands during pre-processing. It comprised of 5211 labelled samples with a total of 13 types of land-cover classes. Table 1 provides details of land-cover classes. The BOT dataset[33], shown in Figure 4 b, was acquired from Botswana on 31 May 2001. It comprises 242 spectral bands and was reduced to 145 spectral bands after removal of noisy and non-calibrated bands during pre-processing. It comprised of 3248 labelled samples with a total of 14 types of land-cover classes. Table 2 provides the details of land-cover classes.

Experimental set-up

The aim of the present study was to classify HSI using the AL technique. Further, the focus was to minimize the computational cost that was gained by utilizing ELM. The ELM with sigmoid activation function having 60 hidden nodes has been used here, as this combination provided better results in case of KSC and BOT datasets in our earlier studies[16,42]. Further, we studied the combination of ELM classifier with EQB query strategy and also compared it with RS and MV query strategies. The parameter k of EQB was set to 4 and every kth bootstrap contained 60% of C_L (ref. 20). In case of the MV strategy, the five views were generated as defined by Di and Crawford[11] with band indices 1–11, 12–31, 32–96, 97–130 and 131–176 for the KSC dataset. Similarly, the five views were generated with band indice: 1–25, 26–61, 62–79, 80–110 and 111–145 for the BOT dataset. The initial training set (C_L) randomly picked five samples from each class of the training set. Hence, the initial training set consisted of 65 samples (5 samples × 13 classes) for KSC and 70 samples (5 samples × 14 classes) for BOT. The total number of iterations was fixed as 200 with ten-fold cross-validation and a batch of three samples with the most uncertainty was selected by applying Q in each iteration. The experiments were performed for ten trials with randomly chosen C_L, C_U and C_T from the original dataset in each trial to reduce the effect of randomness. The overall accuracy was determined by averaging the accuracy of ten trials. Further, computation time was calculated by averaging the computation time of ten trials. Table 3 presents the characteristics and experimental parameters of KSC and BOT datasets.

A comprehensive comparison of the six systems formed by criss-cross combinations of two models (ELM, SVM) and three query strategies (RS, MV, EQB) resulted in six systems: (i) ELM–RS; (ii) ELM–MV; (iii) ELM–EQB (iv) SVM–RS; (v) SVM–MV and (vi) SVM–EQB. Experiments were conducted using cpu@3.40 GHz, i3-4130 and run on Matlab-2016. Active learning toolbox (ALTB)[48] consisting of multiclass SVM and implemented using torch 3 library, was used in the proposed AL technique.

Results and performance analysis

The performance analysis can be depicted-based on two parameters: (i) computation time and (ii) overall accuracy versus samples in the training set. The performance analysis was carried out for all six combinations: Figure 5
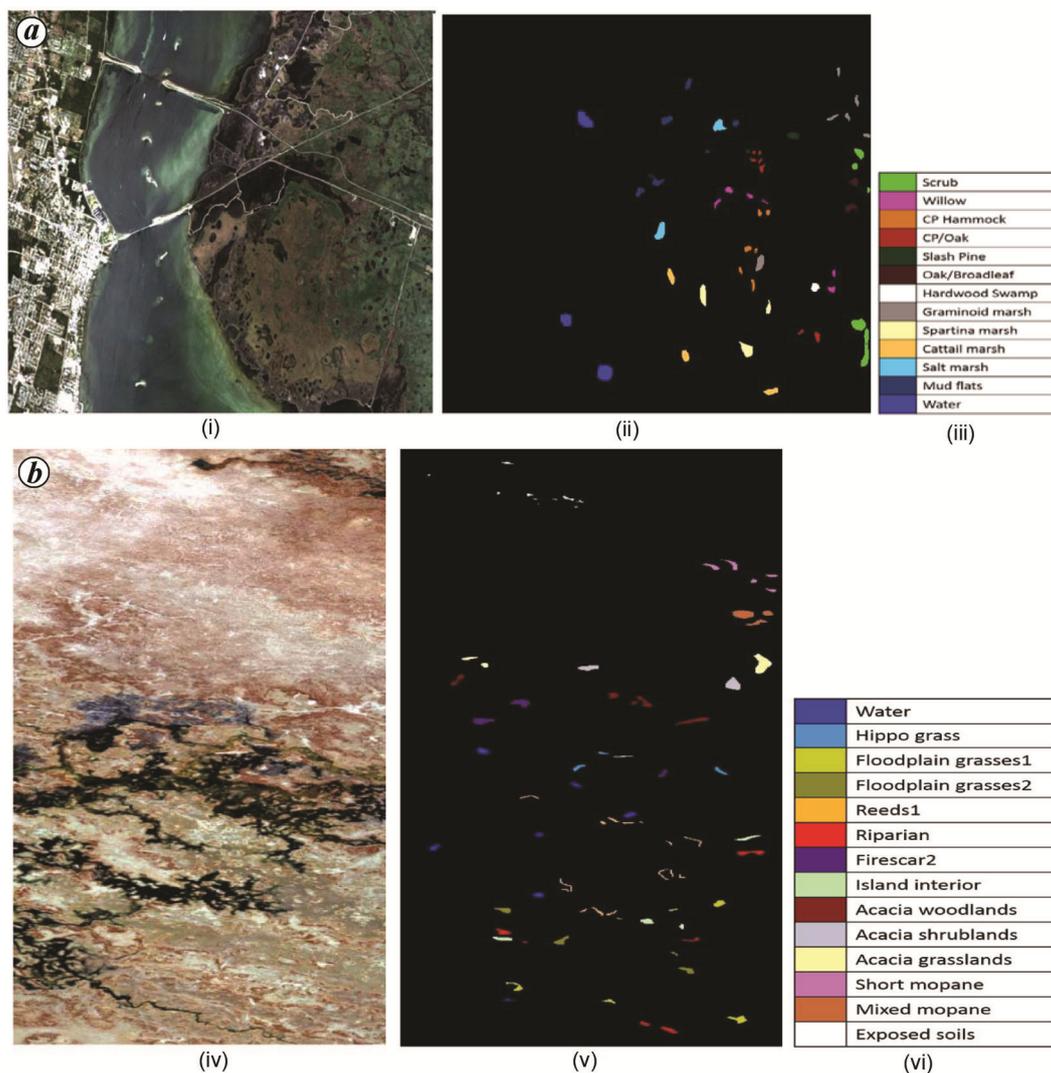
Figure 4. Hyperspectral image (HIS) dataset. *a*, Kennedy Space Centre (KSC) dataset: (i) RBG image, (ii) ground-truth image and (iii) class label image. *b*, Botswana (BOT) dataset: (iv) RBG image, (v) ground-truth image and (vi) class label image.

Table 1. Land-cover classes and the number of samples in Kennedy Space Centre (KSC)

Land-cover class	No. of samples
Water	927
Willow	243
Cattail Marsh	404
CP/oak	252
Slash pine	161
Spartina marsh	520
Hard-wood swamp	105
Graminoid marsh	431
Oak/broadleaf	229
CP hammock	256
Salt marsh	419
Mud flats	503
Scrub	761
Total no. of samples	5211

Table 2. Land-cover classes and the number of samples in Botswana (BOT)

Land-cover classes	No. of samples
Water	270
Hippo grass	101
Floodplain grasses 1	251
Floodplain grasses 2	215
Reeds 1	269
Riparian	269
Firescar 2	259
Island interior	203
Short mopane	181
Acacia shrub-lands	248
Acacia grasslands	305
Acacia woodlands	314
Mixed mopane	268
Exposed soils	95
Total no. of samples	3248

Table 3. Characteristics and experimental parameters of KSC and BOT datasets

Characteristics	KSC	BOT
No. of classes	13	14
No. of samples	5211	3248
Initial no. of samples under each class in C_L	5	5
Initial no. of training sets (C_L)	65	70
Batch size (P) in every iteration	3	3
No. of iterations	200	200
No. of samples in C_L after 200 iterations	665	670
No. of samples in C_U after 200 iterations	1940	954
No. of samples in the test set (C_T)	2606	1624
No. of committees in EQB (k)	4	4
Size of band ranges for multiview	1–11, 12–31, 32–96, 97–130, 131–176	1–25, 26–61, 62–79, 80–110, 111–145

Table 4. Classification accuracy of six systems for KSC and BOT

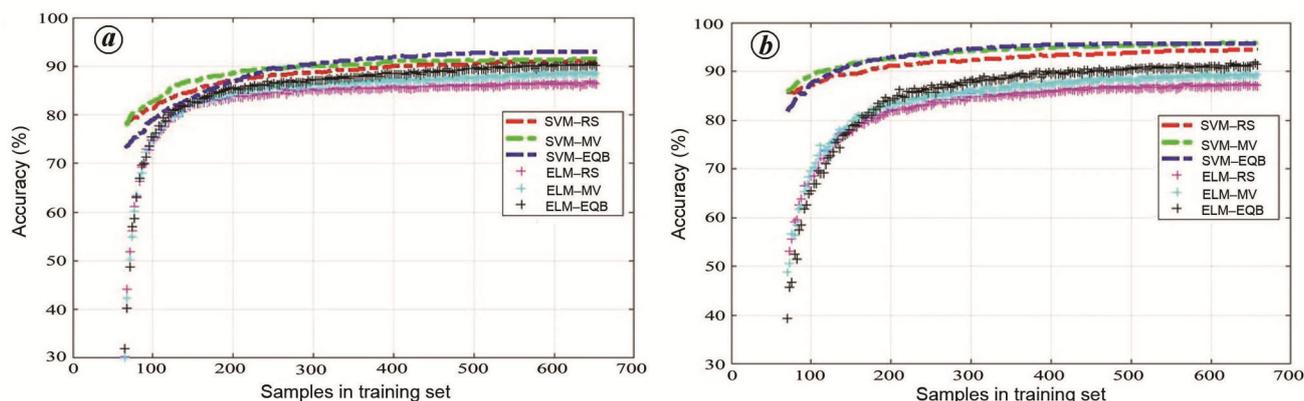|	Classification accuracy	
Active learning (AL) approaches	KSC	BOT
Extreme learning machine–random sampling (ELM–RS)	86.39	87.21
ELM–multiview (MV)	88.60	89.41
ELM–entropy query by bagging (EQB)	90.18	91.31
Support vector machine (SVM)–RS	91.18	94.55
SVM–MV	91.50	95.91
SVM–EQB	92.87	95.93

Figure 5. Classification accuracy versus number of samples in the training set: (***a***) KSC and (***b***) BOT.

shows the learning curves of overall accuracy versus samples in the training set, which highlights the improvement in classification accuracy with increase in the number of samples in the training set. At each iteration, three uncertainty samples were added to the training samples and hence, the number of samples in the training set increased with each iteration. With regard to classification accuracy (Figure 5), The EQB query strategy was better than the MV and RS query strategies for both datasets in all six systems, irrespective of the classifier. Table 4 shows the classification accuracy obtained by the averaging of 10 trials after 200 iterations for the two datasets in all six systems. Table 4 demonstrates that the classification accuracy obtained using ELM-based AL models is slightly less compared to SVM-based AL model for both datasets, irrespective of the query strategy. However, the objective of this study is to achieve the desired classification accuracy with reduced computation time. Figure 6 shows the computation time for all six systems and both datasets. It shows that though the classification accuracy obtained using ELM-based AL model is slightly less compared to SVM-based AL model, computation time in the former is significantly less compared to the latter, irrespective of the query strategy. Figure 5 shows the superior performance of the EQB query strategy and Figure 6 demonstrates that ELM-based AL model outperforms with regard to computation time, while achieving the desired classification accuracy. The above two observations show that the ELM–EQB technique can be a better option in HSI classification, where adequate classification accuracy can be achieved with significantly less computation time.

Table 5 highlights the advantages of the present system, particularly on KSC and BOT. This benchmarking

Table 5. Benchmarking of the present system compared to AL-based classification for KSC and BOT

Author	Dataset	No. of pixels	No. of classes	No. of bands	Query function classifier	Classifier	Initial no. of training samples	Batch size at each iteration	No. of iterations	Classification accuracy (%)	Computation time (sec)
Tuia et al.[20]	KSC	5211	13	176	EQB	SVM	200 pixels	30	70	93.36	--
Demir et al.[22]	KSC	5211	13	176	MCLU–ECBD	SVM	4% pixels/class	40	16	92.82	240
Di et al.[11]	KSC	3784	10	176	MV–AMD	SVM	3 pixels/class	1	870	94.50	--
	BOT	3014	9	145				1	873	97.50	
Wan et al.[33]	KSC	5211	13	176	CASSL	SVM	10 pixels/class	10	87	95.63	1429.6
	BOT	3248	14	145				10	86	95.86	
Pasolli et al.[23]	KSC	5211	13	176	Active-metric learning	SVM	5 pixels/class	10	40	96.20	–
Wang et al.[49]	KSC	5211	13	176	Semi-supervised AL	SVM	10 pixels/class	20	43	93.47	–
	BOT	3248	14	145				20	43	97.03	
Patra et al.[34]	KSC	5211	13	176	Spectra-Spatial Multicriteria AL	SVM	3 pixels/class	20	10	99.71	417
Pradhan et al.[42]	KSC	3784	10	176	MV–AMD	ELM	5 pixels/class	3	300	90.08	38.89
	BOT	2453	9	145				3	300	91.20	27.18
Jamshidpour et al.[17]	KSC	5211	13	176	MV-based genetic algorithm	SVM	50 pixels	5	40	87.21	523.86
The present system	KSC	5211	13	176	EQB	ELM	5 pixels/class	3	200	90.18	119.88
	BOT	3248	14	145				3	200	91.31	88.87

MCLU–ECBD, Multiclass level uncertainty–enhanced clustering-based diversity; MV–AMD, Multiview-based adaptive maximum disagreement; CASSL, Collaborative active and semi-supervised learning.
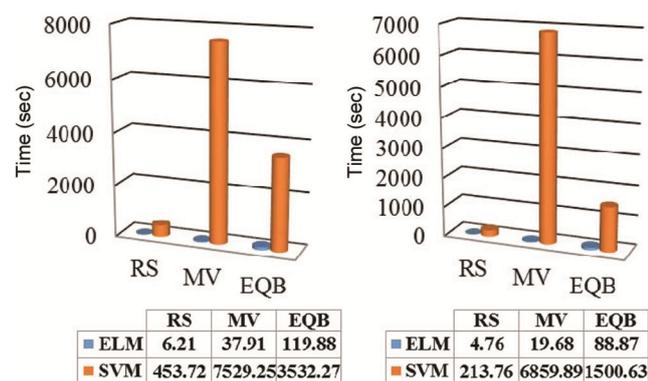
Figure 6. Computation time of all six systems (ELM–random sampling (RS), ELM–multiview (MV), ELM–EQB, SVM–RS, SVM–MV and the SVM–EQB) on two datasets: (*a*) KSC and (*b*) BOT.

was carried out based on different significant AL parameters. As observed in Table 5, all the existing studies have used SVM classifier for AL framework, except for Pradhan *et al.*[16,42], who have studied ELM classifier for the AL framework. Using SVM as a classifier for the AL framework, Tuia *et al.*[20] studied the EQB query strategy to achieve 93.36% classification accuracy in 70 iterations on KSC. Demir *et al.*[23] studied the multiclass level uncertainty-enhanced clustering-based diversity (MCLU–ECBD) to achieve 92.82% accuracy in 16 iterations with computational time of 240 sec on KSC. Di and Crawford[11] studied MV–AMD query strategy to achieve 94.50% and 97.50% accuracy in 870 and 873 iterations on KSC and BOT respectively. Wan *et al.*[33] studied collaborative active and semi-supervised learning (CASSL) to achieve 95.63% and 95.86% accuracy in 86 and 87 iterations on KSC and BOT respectively, and have only depicted computational time of 1429.6 sec for KSC. Pasolli *et al.*[23] used active-metric learning approach to achieve 96.20% accuracy in 40 iterations on KSC. Wang *et al.*[49] have explored semi-supervised AL to achieve 93.47% accuracy on KSC and 97.03% accuracy on BOT in 43 iterations. The spectral–spatial multicriteria AL approach was studied by Patra *et al.*[34] on KSC, to achieve a classification accuracy of 99.71% in ten iterations with a computational time of 417 sec. In contrast, Pradhan *et al.*[16,42] studied the ELM–AL framework with MV–AMD query strategy to achieve a classification accuracy of 90.08% on KSC and 91.20% on BOT in 300 iterations with computation time of 38.89 sec and 27.18 sec respectively. In the present study, ELM–AL with the EQB query strategy has been explored and found to achieve 90.18% accuracy of classification on KSC and 91.31% on BOT in 200 iterations with computational time of 119.88 sec and 88.87 sec respectively. As depicted in Table 5, only a few studies have analysed computation time. However, the present system has an advantage of providing adequate classification accuracy and also significantly reduces computation time.

Conclusion

The present study was undertaken with the objective to evolve an AL approach for classification of HSI with less computation time, while maintaining adequate classification accuracy. With this objective, this study presented an approach of AL to classify HSI using the EQB query strategy in the ELM classifier framework. Further, a comprehensive comparison has been made on six systems formed by criss-cross combinations of three query strategies (RS, MV, EQB) and two classifiers (ELM, SVM). The study showed that adequate classification accuracy can be achieved with significantly less computation time using the ELM-based AL model compared to the state-of-the-art SVM-based AL models, irrespective of the query strategies. In addition, experiments have demonstrated that the EQB query strategy provides better classification accuracy than RS and MV, irrespective of the classifiers used. Therefore, the encouraging outcome of this study is that the ELM–EQB AL approach is a good option for HSI classification. However, there is scope for the proposed model to improve its performance. The spectral information of the image has only been used here. Therefore, performance of the ELM–AL approach can be improved by including spatial information[6,34,41]. Further, we intend to add diversity criteria in the future along with uncertainty criteria in our model to improve its performance[22].

1. Pearlman, J. S., Berry, P. S., Segal, C. C., Shapanski, J., Beiso, D. and Carman, S. L., Hyperion: a space-based imaging spectrometer. *IEEE Trans. Geosci. Remote Sensing*, 2003, **41**(6), 1160–1173.
2. Heiden, U., Heldens, W., Roessner, S., Segl, K., Esch, T. and Mueller, A., Urban structure type characterization using hyperspectral remote sensing and height information. *Lands. Urban Plann.*, 2012, **105**(4), 361–375.
3. Giri, C., Observation and monitoring of mangrove forests using remote sensing: opportunities and challenges. *Remote Sensing*, 2016, **8**(9), 1–8.
4. Sahoo, R. N., Ray, S. S. and Manjunath, K. R., Hyperspectral remote sensing of agriculture. *Curr. Sci.*, 2015, **108**(5), 848–859.
5. Jiang, J. and Tian, G., Analysis of the impact of land use/land cover change on land surface temperature with remote sensing. *Proc. Environ. Sci.*, 2010, **2**, 571–575.
6. Fauvel, M., Tarabalka, Y., Benediktsson, J. A., Chanussot, J. and Tilton, J. C., Advances in spectral–spatial classification of hyperspectral images. *Proc. IEEE*, 2013, **101**(3), 652–675.
7. Camps-Valls, G., Tuia, D., Bruzzone, L. and Benediktsson, J. A., Advances in hyperspectral image classification: earth monitoring with statistical learning methods. *IEEE Signal Process. Mag.*, 2014, **31**(1), 45–54.
8. Kuo, B., Ho, H., Li, C., Hung, C. and Taur, J., A kernel-based feature selection method for SVM with RBF Kernel for hyperspectral image classification. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sensing*, 2014, **7**(1), 317–322.
9. Fan, Y., Zheng, S., Yang, H., Zhang, C. and Su, H., Causality-weighted active learning for abnormal event identification-based on the topic model. *Opt. Eng.*, 2012, **51**(7), 077204-1–077204-12.
10. Han, Y., Li, P., Zhang, Y., Hong, Z., Liu, K. and Wang, J., Combining active learning and transductive support vector machines for sea ice detection. *J. Appl. Remote Sensing*, 2018, **12**(2), 026016.

11. Di, W. and Crawford, M. M., View generation for multiview maximum disagreement based active learning for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sensing*, 2012, **50**(5), 1942–1954.

12. Mitra, P., Murthy, C. A. and Pal, S. K., A probabilistic active support vector learning algorithm. *IEEE Trans. Pattern Anal. Mach. Intel.*, 2004, **26**(3), 413–418.

13. Rajan, S., Ghosh, J. and Crawford, M. M., An active learning approach to hyperspectral data classification. *IEEE Trans. Geosci. Remote Sensing*, 2008, **46**(4), 1231–1242.

14. Tuia, D., Pasolli, E. and Emery, W. J., Using active learning to adapt remote sensing image classifiers. *Remote Sensing Environ.*, 2011, **115**(9), 2232–2242.

15. Crawford, M. M., Tuia, D. and Yang, H. L., Active learning: any value for classification of remotely sensed data. *Proc. IEEE*, 2013, **101**(3), 593–608.

16. Pradhan, M. K., Minz, S. and Shrivastava, V. K., A kernel-based extreme learning machine framework for classification of hyperspectral images using active learning. *J. Indian Soc. Remote Sensing*, 2019, **47**(10), 1693–1705.

17. Jamshidpour, N., Safari, A. and Homayouni, S., Multiview active learning optimization based on genetic algorithm and Gaussian mixture models for hyperspectral data. *IEEE Geosci. Remote Sensing Lett.*, 2019, **17**(1), 172–176.

18. Pradhan, M. K., Minz, S. and Shrivastava, V. K., Fisher discriminant ratio based multiview active learning for the classification of remote sensing images. In 4th International Conference on Recent Advances in Information Technology, Dhanbad, India, 2018, pp. 1–6.

19. Tuia, D., Volpi, M., Copa, L., Kanevski, M. and Munoz-Mari, J., A survey of active learning algorithms for supervised remote sensing image classification. *IEEE J. Sel. Top. Signal Process.*, 2011, **5**(3), 606–617.

20. Tuia, D., Ratle, F., Pacifici, F., Kanevski, M. F. and Emery, W. J., Active learning methods for remote sensing image classification. *IEEE Trans. Geosci. Remote Sensing*, 2009, **47**(7), 2218–2232.

21. Scheffer, T., Decomain, C. and Wrobel, S., Active hidden Markov models for information extraction. In Proceedings of the 4th International Symposium on Intelligent Data Analysis, Cascais, Portugal, 2001, pp. 309–318.

22. Pasolli, E., Melgani, F. and Bazi, Y., Support vector machine active learning through significance space construction. *IEEE Geosci. Remote Sensing Lett.*, 2011, **8**(3), 431–435.

23. Demir, B., Persello, C. and Bruzzone, L., Batch-mode active-learning methods for the interactive classification of remote sensing images. *IEEE Trans. Geosci. Remote Sensing*, 2011, **49**(3), 1014–1031.

24. Li, J., Bioucas-Dias, J. M. and Plaza, A., Semisupervised hyperspectral image segmentation using multinomial logistic regression with active learning. *IEEE Trans. Geosci. Remote Sensing*, 2010, **48**(11), 4085–4098.

25. Li, J., Bioucas-Dias, J. M. and Plaza, A., Hyperspectral image segmentation using a new Bayesian approach with active learning. *IEEE Trans. Geosci. Remote Sensing*, 2011, **49**(10), 3947–3960.

26. Platt, J. C., Probabilistic outputs for support vector machines and comparisions to regulized likelihood methods. In *Adv. Large Margin Classifiers*, 1999, **10**(3), 61–74.

27. MacKay, D. J. C., Information-based objective functions for active data selection. *Neural Comput.*, 1992, **4**, 590–604.

28. Luo, T., Kramer, K., Goldgof, D. B., Hall, L. O., Samson, S., Remsen, A. and Hopkins, T., Active learning to recognize multiple types of plankton. *J. Mach. Learn. Res.*, 2005, **6**, 589–613.

29. Zhou, Y. and Goldman, S., Democratic co-learning. In 16th IEEE International Conference, Tools with Artificial Intelligence, Boca Raton, FL, USA, 2004, pp. 594–602.

30. Abe, N. and Mamitsuka, H., Query learning strategies using boosting and bagging. In Proc. ICML, Madison, WI, USA, 1998, pp. 1–9.

31. Melville, P. and Mooney, R., Diverse ensembles for active learning. In Proceedings of the 21st International Conference on Machine Learning, Banff, Canada, 2004, pp. 584–591.

32. Xu, X., Li, J. and Li, S., Multiview intensity-based active learning for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sensing*, 2018, **56**(2), 669–680.

33. Wan, L., Tang, K., Li, M., Zhong, Y. and Qin, A. K., Collaborative active and semisupervised learning for hyperspectral remote sensing image classification. *IEEE Trans. Geosci. Remote Sensing*, 2015, **53**(5), 2386–2396.

34. Patra, S., Bhardwaj, K. and Bruzzone, L., A spectral–spatial multicriteria active learning technique for hyperspectral image classification. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sensing*, 2017, **10**(12), 5213–5227.

35. Huang, G. B., Zhu, Q. Y. and Siew, C. K., Extreme learning machine: theory and applications. *Neurocomputing*, 2006, **70**(1–3), 489–501.

36. Huang, G. B., Wang, D. H. and Lan, Y., Extreme learning machine: a survey. *Int. J. Mach. Learn. Cybern.*, 2011, **2**(2), 107–122.

37. Huang, G. B., Zhou, H., Ding, X. and Zhang, R., Extreme learning machine for regression and multiclass classification. *IEEE Trans. Syst., Man Cybern.*, 2012, **42**(2), 513–529.

38. Pal, M., Maxwell, A. E. and Warner, T. A., Kernel-based extreme learning machine for remote-sensing image classification. *Remote Sensing Lett.*, 2013, **4**(9), 853–862.

39. Bazi, Y., Alajlan, N., Melgani, F., AlHichri, H., Malek, S. and Yager, R. R., Differential evolution extreme learning machine for the classification of hyperspectral images. *IEEE Geosci. Remote Sensing Lett.*, 2014, **11**(6), 1066–1070.

40. Samat, A., Du, P., Liu, S. , Li, J. and Cheng, L., E^2LMs: ensemble extreme learning machines for hyperspectral image classification. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sensing*, 2014, **7**(4), 1060–1069.

41. Heras, D. B., Argüello, F. and Quesada-Barriuso, P., Exploring ELM based spatial–spectral classification of hyperspectral images. *Int. J. Remote Sensing*, 2014, **35**(2), 401–423.

42. Pradhan, M. K., Minz, S. and Shrivastava, V. K., Fast active learning for hyperspectral image classification using extreme learning machine. *IET Image Process.*, 2018, **13**(4), 549–555.

43. Li, M. and Sethi, I. K., Confidence-based active learning. *IEEE Trans. Pattern Anal. Mach. Intellige.*, 2006, **28**(8), 1251–1261.

44. Seung, H. S., Opper, M. and Sompolinsky, H., Query by committee. In Proceedings of the Fifth Annual Workshop Computational Learning Theory, Pittsburgh, PA, USA, 1992, pp. 287–294.

45. Freund, Y., Seung, H. S., Shamir, E. and Tishby, N., Selective sampling using the query by committee algorithm. *Mach. Learn.*, 1997, **28**(2–3), 133–168.

46. Efron, B., Bootstrap methods: another look at the jackknife. In *Breakthroughs Statistics,* Springer, New York, USA, 1992, pp. 569–593.

47. HSI dataset: KSC and BOT; http://www.ehu.eus/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes (accessed on 22 September 2017).

48. Chang, C. C. and Lin, C. J., LIBSVM: a library for support vector machines. *ACM Trans. Intel. Syst. Techn.*, 2011, **2**(3), 27; http://www.csie.ntu.edu.tw/~cjlin/libsvm (accessed on 22 October 2017).

49. Wang, Z., Du, B., Zhang, L., Zhang, L. and Jia, X., A novel semisupervised active-learning algorithm for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sensing*, 2017, **55**(6), 3071–3083.