





# GPU-Accelerated CatBoost-Forest for Hyperspectral Image Classification Via Parallelized mRMR Ensemble Subspace Feature Selection

Alim Samat , Member, IEEE, Erzhu Li , Peijun Du, Senior Member, IEEE, Sicong Liu , Member, IEEE, and Junshi Xia , Senior Member, IEEE

**Abstract**— In this article, the graphics processing unit (GPU)-accelerated CatBoost (GPU-CatBoost) algorithm for hyperspectral image classification is first introduced and comparatively studied using diverse features. To further foster the classification performance from both accurate and efficient viewpoints, an ensemble version of GPU-CatBoost, the GPU-accelerated CatBoost-Forest (GPU-CatBF) algorithm, is proposed by adopting the parallelized minimum redundancy maximum relevance (mRMR) ensemble (PmRMRE) feature selection (FS) algorithm. To evaluate the performance and suitability of mRMR and PmRMRE, 11 other state-of-the-art FS algorithms are comprehensively investigated. Experimental results on three widely acknowledged hyperspectral benchmarks showed that: 1) GPU-CatBoost is also an advanced ensemble learning (EL) algorithm for hyperspectral image classification using diverse features; 2) mRMR and PmRMRE have advanced properties for highly discriminative features and band selection, and the best results are achieved by PmRMRE in most cases in terms of both the robustness and computational efficiency; and 3) GPU-CatBF always outperforms CatBoost and GPU-CatBoost, while compatible and even better results are reachable without losing much computational efficiency in contrast with other selected decision tree-based EL algorithms.

**Index Terms**—CatBoost, ensemble learning (EL), feature selection (FS), gradient boosted decision tree (GBDT), histogram-based gradient boosted trees (histGBT), hyperspectral image classification, lightGBM, minimum redundancy maximum relevance (mRMR), parallelized mRMR ensemble (PmRMRE).

Manuscript received December 1, 2020; revised December 28, 2020; accepted February 27, 2021. Date of publication March 3, 2021; date of current version March 25, 2021. This work was supported in part by the National Natural Science Foundation of China under Grant 42071424 and Grant 42071324, and in part by the Youth Innovation Promotion Association Foundation of the Chinese Academy of Sciences under Grant 2018476. (Corresponding author: Alim Samat.)

Alim Samat is with the State Key Laboratory of Desert and Oasis Ecology, Xinjiang Institute of Ecology and Geography, CAS, Urumqi 830011, China, with the Research Center for Ecology and Environment of Central Asia, CAS, Urumqi 830011, China, and also with the University of Chinese Academy of Sciences, Beijing 100049, China (e-mail: alim\_smt@ms.xjb.ac.cn).

Erzhu Li is with the Department of Geographical Information Science, Jiangsu Normal University, Xuzhou 221100, China (e-mail: lierzhu2008@126.com).

Peijun Du is with the School of Geography and Ocean Science, Nanjing University, Nanjing 210023, China (e-mail: peijun@nju.edu.cn).

Sicong Liu is with the College of Surveying and Geoinformatics, Tongji University, Shanghai 200092, China (e-mail: sicong.liu@tongji.edu.cn).

Junshi Xia is with the Geoinformatics Unit, RIKEN Center for Advance Intelligence Project, Tokyo 103-0027, Japan (e-mail: junshi.xia@riken.jp).

Digital Object Identifier 10.1109/JSTARS.2021.3063507

## I. INTRODUCTION

LAND cover mapping is one of the main applications of remote sensing (RS) data and is essential for understanding the patterns and driving factors of land cover changes on the earth's surface [1], [2]. Over the past 40 years, large numbers of supervised, unsupervised, and semisupervised shallow and deep classification methods have been developed to map land cover using RS data due to their superior robustness compared to model-based approaches [3]–[6]. Theoretically and more practically, an ideal supervised classifier should be capable of addressing the following challenges.

- 1) The curse of dimensionality, i.e., the Hughes phenomenon [7].
- 2) The nonlinearity of variables [8].
- 3) Small and class-imbalanced training samples [9].
- 4) Feature noise and label noise in both the labeled and unlabeled samples [10].
- 5) Easy implementation and yet computationally efficient [11].

Driven by such challenges, advanced machine learning (ML) algorithms, such as support vector machines (SVMs) [4], artificial neural networks (ANNs) [12], extreme learning machines (ELMs) [13], decision forests (DFs) [14], [15], and deep neural networks (DNNs) [16], [17], have emerged as more accurate and efficient alternatives to conventional model-based approaches, particularly when faced with high-dimensional, complex data spaces, multitemporal, and large-area mapping cases [18], [19]. However, some ML algorithms (e.g., SVMs, ELMs, ANNs, and DNNs) are complicated due to having critical model parameters that need to be tuned first, which is difficult to automate [12]–[17]. Additionally, these algorithms have a tendency to overfit the given data, especially in small-sample training scenarios. Among these algorithms, ensemble learning (EL) algorithms that utilize ensembles of decision trees (DTs) have received increasing interest due to their ease of deployment, fast operation, more accurate classification results, and robustness to noise compared to single classifiers [6], [14], [15], [20]–[22]. In this family of EL, classifiers, including the random forest (RaF), rotation forest (RoF), extremely randomized DTs (ExtraTrees), extreme gradient boosting trees (XGBoost), and deep decision forest (DDF) classifiers, are increasingly being adopted in a variety of classification tasks using airborne and spaceborne

multi/hyperspectral, light detection and ranging (LiDAR), synthetic aperture radar (SAR), and polarimetric SAR (PolSAR) systems [21]–[28].

As a novel and recent modification of ordered gradient boosting (OGB) but with categorical feature support, CatBoost has outperformed existing state-of-the-art algorithms such as gradient boosted DT (GBDT) [29], XGBoost [30], LightGBM [31], and H2O [32] on a diverse set of popular ML tasks on both central processing unit (CPU) and graphics processing unit (GPU) implementations [33]. However, according to our previous work in [34], the superior performance of CatBoost in terms of the classification accuracy was observed in only a few cases with a large value set of boosting iterations, and it was less computationally efficient than adaptive boosting (AdaBoost), GBDT, XGBoost with the classification and regression trees (CART) booster, and LightGBM on CPU implementation. This limitation could become an important challenge in the classification of large volumes of hyperspectral images with high dimensionality, especially in time-critical applications. Thanks to recent advances in high-performance computing techniques, accurate and efficient classification performance can be achieved for adopted classifiers by exploiting specialized devices, such as clusters and distributed computers, multicore CPUs, field-programmable gate arrays (FPGAs), and GPUs in hyperspectral image processing [35], [36]. Specifically, it is possible to greatly accelerate the computational efficiency of a classifier on a GPU-based parallel computing platform by benefiting from its capacity of performing many computationally intensive tasks in parallel [36]–[38]. Once the computational complexity of the adopted classifier is greatly accelerated, it is also possible to further boost the classification accuracy by constructing an EL system [36], [37], [39]. Hence, it is of interest to investigate the performance of the GPU-accelerated CatBoost (GPU-CatBoost) algorithm and its ensemble version in hyperspectral image classification using diverse features.

An EL system usually consists of two key components—a strategy to produce classifiers with high diversity and a rule to combine the results from multiple classifiers. The first key component, which is the cornerstone for constructing an effective EL system, can be achieved by resampling, label switching, feature partitioning, feature selection (FS), feature extraction (FE), model parameter shuffling, and hybridization techniques [5], [13], [40]–[42]. In contrast with other methods, FS- and FE-based techniques are capable of addressing the curse of dimensionality and tasks with high feature-to-instance ratios [43], [44]. In contrast with FS-based EL algorithms, the computational burden brought about by the FE procedure is always a challenge, particularly in the classification of large volumes of hyperspectral images with high dimensionality [15], [25]. Additionally, in the sense of maintaining the statistical and physical meanings of original features, FS is superior in terms of better readability and interpretability. Hence, we selected the FS strategy to construct our proposed GPU-CatBF model.

Theoretically, any FS method can be adopted to produce classifiers with high diversity to construct an EL system. However, practically speaking, a robust and highly efficient FS method

is always the best option. Dimensionality reduction via FS, including band selection, is one of the most popular techniques to remove noise and redundant features, improve the learning performance, reduce the computational cost, build models with better generalizability, and decrease the amount of storage required in the context of hyperspectral image processing [45]–[47]. Therefore, numerous FS algorithms have been introduced and proposed in this field in the past few decades.

Based on whether a labeled training set is available, FS algorithms can be grouped into supervised, unsupervised, and semisupervised algorithms [48]–[50]. For the classification problem, FS aims to select highly discriminate features that are capable of discriminating samples belonging to different classes. In this regard, the supervised FS algorithm works better than unsupervised and semisupervised algorithms when sufficiently labeled samples are available. Based on the relationship between an FS algorithm and the inductive learning method used to infer a model, supervised FS algorithms can further be broadly categorized into three types—filter-, wrapper-, and embedded-based methods [49]. Compared with wrapper-based methods, which use a single learner as a black box to evaluate the subsets of features according to their predictive power, and embedded-based methods, which perform FS in the training process and are usually specific to a given learning algorithm, the filter-based method, which selects the subset of features as a preprocessing step independent of the induced algorithm, is advantageous for its low computational cost and good generalization ability [51]. Bing is one of the most powerful filter-based FS methods among the ML community, as shown by its high citation count (more than 8000). The minimum redundancy maximum relevance (mRMR) [52] algorithm has been extensively studied in the fields of DNA microarray data classification [53], protein classification [54], gene expression [55], water resource system management [56], 3-D facial expression recognition [57], lung cancer detection [58], real-time static voltage stability assessment [59], and many others. However, only a few works have introduced and investigated the performance of mRMR for band selection from multispectral and hyperspectral images [60]–[63].

To further boost the performance on maximal relevance and minimal redundant FS according to the mutual information (MI)-based maximal statistical dependency criterion, mRMRE is an extension of mRMR by using an ensemble technique [59]. However, mRMR is a centralized method, and it scales quadratically with the number of features and grows linearly with respect to the sample size [52], [65]. As a result, the computationally expensive limitation of mRMR will be inherited and further enhanced in mRMRE and become a serious challenge in high dimensionality and large-sample scenarios. To tackle this limitation, proposals have been made on the acceleration of mRMR and mRMRE using efficient parallelization techniques [65]–[67]. To the best of our knowledge, parallelized mRMRE (PmRMRE) has not yet been studied for FS-based hyperspectral image classification, either on its CPU- or GPU-based implementations. Therefore, in this work, to satisfy the need for a robust and highly efficient FS algorithm from the construction of an FS-based EL system, PmRMRE is selected to construct the

subspace ensemble version of GPU-CatBoost, the GPU-CatBF, for hyperspectral image classification.

For the second key component of an effective EL system, popular fusion methods, including majority voting, weighted majority voting, error pruning, meta fusion, Bayesian fusion, fuzzy integral, D-S evidence theory, consensual theory, Borda count, and algebraic rule-based methods, can be used to combine results from multiple classifiers [41], [68]. However, on the one hand, advanced but complex fusion methods are not suitable for large data with large ensemble sizes from a computationally efficient point of view; on the other hand, simple voting-based methods could limit or even degrade algorithm performance in small ensembles and in large ensembles with lower classifier diversity [26]. To overcome such limitations, a metaensemble criterion that might yield the best solution was adopted.

The main contributions of this article are summarized as follows.

- 1) GPU-CatBoost was introduced and comparatively investigated for hyperspectral image classification using diverse features.
- 2) mRMR and PmRMRE were introduced and comparatively evaluated for discriminative subspace FS-based hyperspectral image classification.
- 3) To improve the classification performance, a new ensemble version of GPU-CatBoost, GPU-CatBF, was proposed by combining multiple GPU-CatBoost models trained on discriminative subspace features from PmRMRE.

## II. RELATED WORKS

### A. CatBoost

Many solid theoretical and empirical results indicate that gradient boosting is a powerful ML method, especially for dealing with noisy data, heterogeneous features, and complex dependencies [29]–[33]. However, similar to standard boosting, classical gradient boosting also suffers from overfitting caused by the prediction shift in the learned model, also known as a special kind of target leakage [88]. Furthermore, categorical features with discrete sets of values that are not necessarily comparable to each other cannot be directly handled by binary trees. A common solution for using categorical features in gradient boosting is converting them to numerical features. In this regard, one-hot encoding (OHE), gradient statics (GS), target statics (TS), greedy TS, holdout TS, and leave-one-out TS solutions have been identified. Unfortunately, this transformation procedure can also cause target leakage and prediction shifts [88]. Hence, to avoid both issues of overfitting and target leakage caused by gradient boosting categorical feature transformation, CatBoost was proposed as a combination of OGB and ordered TS [33], [88].

Let  $D = \{(\mathbf{X}_i, Y_i)\}_i^n$  represent a given set of  $n$  samples, where  $\mathbf{X}_i = (x_{i,1}, \dots, x_{i,d})$  is a vector of  $d$  features (some numerical and some categorical) and  $Y_i \in \mathbb{R}$  is a label value. CatBoost substitutes the categorical feature  $x_{\sigma_p,k}$  with

$$x_{\sigma_p,k} = \frac{\sum_{j=1}^{p-1} [x_{\sigma_j,k} = x_{\sigma_p,k}] Y_{\sigma_j} + a \cdot P}{\sum_{j=1}^{p-1} [x_{\sigma_j,k} = x_{\sigma_p,k}] + a}, \quad k \in (1, d) \quad (1)$$

where  $\sigma = (\sigma_1, \dots, \sigma_s)$  is the number of  $s$  random permutations of the dataset and  $a$  is the weight of the prior value  $P$ . Then, CatBoost can be built by following the pseudocode steps in Algorithm 1. where  $I$  is the number of boosting iterations,  $L$

---

### Algorithm 1: Pseudocode for CatBoost [88]

---

**Inputs:**  $\{(\mathbf{X}_i, Y_i)\}_i^n, I, a, L, s, Mode$

**Process:**

- 1)  $\sigma_r \leftarrow$  random permutation of  $[1, n], r = 0, \dots, s;$
- 2)  $M_0(i) \leftarrow 0$  for  $i = 0, \dots, n;$
- 3) **If** Mode = Plain **then**
- 4)  $M_r(i) \leftarrow 0$  for  $r = 1, \dots, s, i : \sigma_r(i) \leq 2^{j+1};$
- 5) **If** Mode = Ordered **then**
- 6) **For**  $j \leftarrow 1$  **to**  $\lceil \log 2n \rceil$  **do**
- 7)  $M_{r,j}(i) \leftarrow 0$  for  $r = 1, \dots, s, i = 1, \dots, 2^{j+1}$
- 8) **For**  $t \leftarrow 1$  **to**  $I$  **do**
- 9)  $T_t, \{M_r\}_{r=1}^s \leftarrow$   
 $Tree(\{M_r\}_{r=1}^s, \{(x_i, y_i)\}_{n=1}^n, a, L, \{\sigma_i\}_{i=1}^s, Mode);$
- 10)  $Leaf_0(i) \leftarrow GetLeaf(x_i, T_t, \sigma_0)$  for  $i = 1, \dots, n;$
- 11)  $g_0 \leftarrow Gradient(L, M_0, y);$
- 12) **Foreach** Leaf  $j$  in  $T_t$  **do**
- 13)  $b_j^t = -avg(g_0(i) \text{ for } i : Leaf_0(i) = j);$
- 14)  $M_0(i) \leftarrow M_0(i) + ab_j^t \Lambda_{Leaf_0(i)}$  for  $i = 1, \dots, n$

**Return:**  $F(x) = \sum_{t=1}^I \sum_j ab_j^t \Lambda_{\{GetLeaf(x, T_t, Mode)=j\}}$

---

is the loss function,  $M_r(i)$  is the support model from the  $r$ th permutation using instance  $x_i$ , and Mode is the boosting modes of plain and ordered. The former mode is the standard GBDT algorithm with inbuilt ordered TS. Due to limited space, [33] and [88] are recommended to readers interested in more detailed algorithmic descriptions. Both CPU and GPU implementation of CatBoost.<sup>1</sup>

### B. Minimum Redundancy Maximum Relevance

In pattern recognition applications, the definition of optional characterization often means the minimum classification error. In an unsupervised case where the classifiers are not specified, minimal error requires the maximal statistical dependency of target class  $c$  on the data distribution in the selected subspace  $R^m$ . However, it is often difficult to obtain an accurate estimation of the maximal dependency for multivariate density, which often involves ill-posed problems. In addition, the computational complexity drawback of maximal dependency is the most pronounced problem, not only for continuous feature variables but also for discrete and categorical features. Alternatively, the maximal relevance, which is usually characterized in terms of correlation or MI, can be used to realize maximal dependency efficiently [53].

In terms of the MI, the maximal dependency criterion tries to find a feature set of  $S$  with  $m$  features, which jointly has the largest dependency on target class  $c$  [53]

$$\max D(S, c), D = MI(\{x_i, i = 1, 2, \dots, m\}; c) \quad (2)$$

<sup>1</sup>Online. [Available]: <https://catboost.ai/>

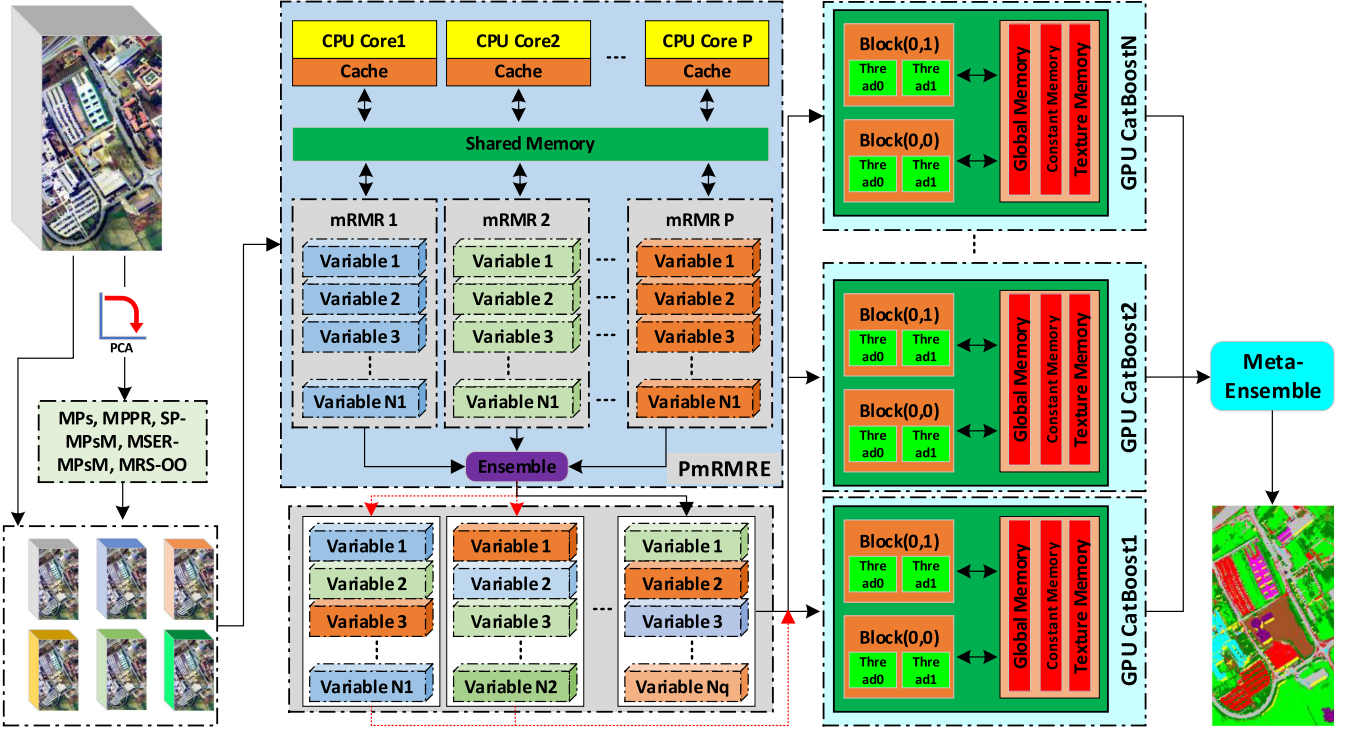


Fig. 1. Technical flowchart for the GPU-CatBF algorithm.

where  $MI(x_i; c)$  represents the MI between feature  $x_i$  and class  $c$ . To approximate (2), the maximal relevance is measured by the mean value of all the MI values between the individual features and target class  $c$  as follows:

$$\max D(S, c), D = \frac{1}{|S|} \sum_{x_i \in S} MI(x_i; c). \quad (3)$$

However, features that are selected according to (3) could have rich redundancy, and the dependency among features could be large. As a result, the representative class discriminative power would not change much if some of the features were highly dependent on others. To solve these issues, the minimal redundancy condition can be adopted for mutually exclusive features by minimizing the following [54]:

$$\min R(S), R = \frac{1}{|S|^2} \sum_{x_i, x_j \in S} MI(x_i, x_j). \quad (4)$$

In practice, maximum relevance and minimum redundancy cannot always be achieved simultaneously. An optimization criterion that combines the above two constraints into a single constraint is called mRMR [54]

$$\begin{aligned} & \max \Phi(D, R), \Phi = D - R \\ & = \max \left[ \frac{1}{|S|} \sum_{x_i \in S} MI(x_i; c) \right] - \frac{\frac{1}{|S|^2} \sum_{x_i, x_j \in S} MI(x_i, x_j)}{\min(H(x_i), H(x_j))} \end{aligned} \quad (5)$$

where  $H(x_i)$  and  $H(x_j)$  are the entropy of the  $i$ th and  $j$ th features, respectively.

### III. PROPOSED METHOD

Diversity is the cornerstone of constructing an effective EL system, and the underlying rule of thumb theory to this concept is that diversified classifiers lead to uncorrelated errors, which in turn improve the classification accuracy. Although many diversification techniques, as mentioned above, are available, FS-based techniques are not only capable of addressing the curse of dimensionality and high feature-to-instance ratio tasks but are also superior in terms of their computational efficiency and better feature readability and interpretability. Furthermore, the feature subset selection algorithm not only takes the performance of the ensemble into account but also directly supports the diversity of subsets of features. Additionally, from constructing an accurate and efficient EL system point of view, a robust and highly efficient FS algorithm is the best practical option. Therefore, the GPU-CatBF algorithm is proposed by utilizing the GPU-CatBoost and PmRMRE algorithm for subset FS, as presented in Fig. 1.

Indeed, a robust and highly efficient FS algorithm is an ideal choice, but the performance of an EL system constructed simultaneously using a robust FS algorithm (e.g., PmRMRE) and a classifier (e.g., GPU-CatBoost) can be limited due to a lack of diversity. For example, it is highly possible that the advanced FS algorithm PmRMRE could return very similar and even exactly the same feature subsets from two independent runs. To overcome this limitation, an incremental FS strategy was adopted in the subspace FS phase, and a metafusion criterion that might be capable of yielding the best results was adopted in the ensemble phase. Finally, the proposed GPU-CatBF algorithm can be built by following the algorithmic steps described in Algorithm 2.

**Algorithm 2:** Algorithmic Code Steps for CatBoost-Forest

---

**Inputs:** Dataset  $D = \{(\mathbf{X}_i, Y_i)\}_i^n$ ,  
 $\mathbf{X}_i = (x_{i,1}, \dots, x_{i,d})$ ;  
 Base learning algorithm  $\zeta_{\text{GPU-CatBoost}}$ ;  
 Feature selector  $\bar{\lambda}_{\text{PmRMRE}}$ ;  
 Incremental FS range  $[\ell_{\min}, \ell_{\max}]$ ,  $\ell_{\max} \leq d$ ;  
 Incremental step  $v$ ;  
**Process:**  
 1) **for**  $t = \ell_{\min}$  **to**  $\ell_{\max}$  **by** step  $v$ ;  
 2)  $F_t = \bar{\lambda}_{\text{PmRMRE}}(D, t)$ ;  
 3)  $D_t = \text{Map}_{F_t}(D)$ ,  $D_t \in D$ ;  
 4)  $\hat{h}_t = \zeta_{\text{GPU-CatBoost}}(D_t)$ ;  
 5)  $\varepsilon_t = \frac{1}{n} \sum_{i=1}^n \hat{h}_t(x_i) \neq y_i$ ;  
**end**  
**Return:**  $H(x) = \arg \max \left\{ \sum_{t=1}^{\ell_{\max} - \ell_{\min} + 1} \hat{h}_t(x), \hat{h}^* \right\}$

---

where  $F_t$  is a set of features selected by PmRMRE at round  $t$ ,  $D_t$  is the new data keeping only the features in  $F_t$ ,  $\hat{h}_t$  is trained GPU-CatBoost learner using  $D_t$ ,  $\varepsilon_t$  is the error of  $\hat{h}_t$ ,  $\hat{h}^*$  is the learner with the lowest classification error, and  $H(x)$  is the final decision function for CatBoost-Forest.

## IV. DATASETS AND SETUP

## A. Datasets

To evaluate the performances of the considered methods, three hyperspectral benchmark datasets, i.e., the Pavia University, GRSS-DFC2013 Houston, and GRSS-DFC2018 Houston datasets, are utilized in our experiments.

1) *Pavia University*: This hyperspectral image was acquired with a reflective optics system imaging spectrometer (ROSIS) optical sensor, which provides 115 bands with spectral range coverage ranging from 0.43 to 0.86  $\mu\text{m}$ . The geometric resolution is 1.3 m. The image shown in Fig. 1(a) was captured over the Engineering School, University of Pavia, Pavia, Italy. It has  $610 \times 340$  pixels with 103 spectral channels (a few original bands are very noisy and were discarded immediately after data acquisition). The validation data refer to nine land cover classes; Table I shows the details about the number of samples and the legend.

2) *GRSS-DFC2013 Houston*: This hyperspectral image was acquired at a spatial resolution of 2.5 m by the NSF-funded Center for Airborne Laser Mapping (NCALM) over the University of Houston campus and the neighboring urban area on June 23, 2012. It is  $349 \times 1905$  pixels with 144 spectral bands in the spectral range between 380 and 1050 nm. The 15 classes of interest selected by the Data Fusion Technical Committee (DFTC) of the IEEE Geoscience and Remote Sensing Society (GRSS) are reported in Table I with the corresponding number of samples for both the training and validation sets.

3) *GRSS-DFC2018 Houston*: This hyperspectral image was collected by the NCALM at the University of Houston on February 16, 2017, between 16:31 and 18:18 GMT using an ITRES compact airborne spectrographic imager (CASI)-1500

sensor covering a 380–1050 nm spectral range with 48 bands at a 1-m ground sampling distance (GSD). This data cube has been orthorectified and radiometrically calibrated to spectral radiance units (milli-SRU). The data were distributed in radiance, and the image size was  $4172 \times 1202$  pixels. The 20 classes of interest by the DFTC of the GRSS are reported in Table I. The corresponding number of samples for both the training and validation sets account for 1% and 99% of the total ground-truth samples, respectively, from Fig. 2(h).

## B. Experiment Setup

To generate diverse features, shallow but advanced spectral-spatial and object FE methods, including extended morphological profiles (EMPs) [69], extended morphological profiles with partial reconstruction (EMPPR) [70], extended maximally stable extremal region guided morphogenesis profiles with mean pixel values within regions (EMSER-MPsM) [25], [26], superpixel guided morphological profiles with superpixel means (SP-MPsM) [71], and multiresolution segmentation (MRS)-based object-oriented profiles (MRS-OO) [72], are selected. Specifically, all of these methods are applied to the first few (typically three) features after principal component analysis (PCA) is executed on the original hyperspectral datasets.

To analyze the performance of the introduced GPU-CatBoost (Cls7) and the proposed GPU-CatBF algorithm (Cls12), popular DT-based EL algorithms, including histogram-based gradient boosting trees (HistGBT, Cls1) [73], random forest (RaF, Cls2) [74], ExtraTrees (Cls3) [75], multiclass AdaBoost (Cls4) [76], GBDT (Cls5) [29], CatBoost (Cls6) [33], XGBoost with CART booster (XGB-CART, Cls8) [30], XGBoost with RaF booster (XGB-RaF, Cls9), GPU-accelerated XGB-RaF (GPU-XGB-RaF, Cls10) [30], [26], and light gradient boosting machine (LiGBM, Cls11) [31], are applied to the three previously described datasets.

To evaluate the performances of the mRMR and PmRMRE algorithms, popular FS algorithms, namely, the relief-based FS (ReliefF) [77], correlation-based FS (CFS) [78], joint MI (JMI) [79], double input symmetrical relevance (DISR) [80], interaction capping (ICAP) [81], Gini index (GiniI) [82], Fisher score (FishS) [83], MI maximization (MIM) [84], conditional MI maximization (CMIM) [85], conditional infomax FE (CIFE) [86], and trace ratio (TraceR) [87] methods, are selected in the experiments.

Finally, the classification overall accuracy (OA), kappa ( $\kappa$ ) statistic, and code running time in seconds were used to evaluate the classification performances of the considered classifiers and FS algorithms. All the experiments were conducted using Python 3.7.8 on a 64-bit Windows 10 system with an Intel Core i7-7820X 3.60-GHz CPU and 128 GB RAM, and with an NVIDIA Quadro P4000 card with CUDA toolkit version 10.2.

## V. EXPERIMENTAL RESULTS AND ANALYSIS

## A. Evaluation of mRMR and PmRMRE

To evaluate the performance of mRMR and PmRMRE for FS on hyperspectral images with diverse features, we first present the OA values of an RaF model (ensemble size = 200, other

TABLE I  
CLASS NAMES, COLOR CODES, AND SAMPLE DETAILS FOR THE CONSIDERED DATASETS

Dataset #	Class ID	Name & color	Test	Training	Dataset #	Class ID	Name & color	Test	Training
1)	1	Asphalt	6631	548	2)	14	Tennis Court	181	247
	2	Meadows	18,649	540		15	Running Track	187	473
	3	Gravel	2099	392	3)	1	Healthy grass	38,805	391
	4	Trees	3064	524		2	Stressed grass	129,878	130
	5	Metal	1345	265		3	Artificial turf	2463	273
	6	Bare soil	5029	532		4	Evergreen trees	53,779	543
	7	Bitumen	1330	375		5	Deciduous trees	19,971	201
	8	Bricks	3682	514		6	Bare earth	17,884	180
	9	Shadows	947	231		7	Water	958	106
2)	1	Healthy grass	198	1053		8	Residential	158,837	158
	2	Stressed grass	190	1064		9	Nonresidential	893,875	894
	3	Synthetic grass	192	505		10	Roads	183,100	183
	4	Trees	188	1056		11	Sidewalks	135,899	136
	5	Soil	186	1056		12	Crossways	5999	60
	6	Water	182	143		13	Major thoroughfares	185,253	185
	7	Residential	196	1072	14	Highways	39,044	394	
	8	Commercial	191	1046	15	Railways	27,471	277	
	9	Road	193	1053	16	Paved parking lots	45,473	459	
	10	Highway	191	1036	17	Unpaved parking lots	529	58	
	11	Railway	181	1050	18	Cars	26,027	262	
	12	Parking Lot 1	192	1041	19	Trains	21,265	214	
	13	Parking Lot 2	184	285	20	Stadium seats	27,024	272	

parameters set to the default values) with an increasing number of selected features using the considered FS algorithms on the experimental datasets. For a more objective performance evaluation, the mean OA values from ten independent runs of the test were used to draw graphs.

According to the results in Fig. 3, it can be easily noticed that the performance of mRMR can be further boosted by the ensemble, which is in accordance with the results in [59], [65]–[67]; see the “x”-marked lines in light blue for PmRMRE and the “x”-marked lines in cyan for mRMR. Moreover, compared with all the other adopted FS algorithms, PmRMRE shows the fastest convergence speed in all cases of using the original spectral, EMPs, EMPPR, EMSER-MPsM, SP-MPsM, and MRS-OO features. Specifically, see the results shown by using EMPs [see Fig. 3(b)] and EMSER-MPsM [see Fig. 3(d)] features from the Pavia University data, the EMPs [see Fig. 3(g)] features from the GRSS-DFC2013 Houston data, and the raw [see Fig. 3(k)], EMPs [see Fig. 3(l)], and EMPPR [see Fig. 3(m)] features from the GRSS-DFC2018 Houston data.

Advanced performance of mRMR is also true for hyperspectral image FS, as shown by the “x”-marked lines in cyan. For instance, a faster convergence speed than the ReliefF, CFS, TraceR, FishS, MIM, and CIFE algorithms can be observed for mRMR on the original spectral features from the Pavia University data [see Fig. 3(a)]. The secondary fast convergence speeds from mRMR can be observed on the original spectral features from the GRSS-DFC2018 Houston data [see Fig. 3(k)]. In addition, obviously faster convergence speeds than the ReliefF, GiniI, and TraceR algorithms can be observed for mRMR by using the EMPs [see Fig. 3(g) and (l)], EMPPR [see Fig. 3(h)

and (m)], and MRS-OO [see Fig. 3(i) and (n)] features from the GRSS-DFC2013 Houston and GRSS-DFC2013 Houston datasets, respectively. Nevertheless, MIM and CMIM could be alternative choices to mRMR when dealing with highly correlated features with high dimensionality, such as EMPs and EMPPR features from the test images, according to results shown by the “x”-marked lines in pink and brown shown in Fig. 3(g), (h), (l), and (m).

Aside from the robustness, the computational efficiency of an FS algorithm is another key factor that needs to be considered in practice. Hence, in Fig. 4, we present the CPU-based computational time costs for the considered FS algorithms using diverse features from the experimental datasets.

According to the results, we can clearly see that the highest computational efficiency is achieved by PmRMRE on all the considered features from the considered datasets, while FishS and TraceR achieved the second best efficiency, and the lowest computational efficiency is attained by CFS, especially using the MRS-OO features with the highest dimensionality (see the results shown by green bars). Similar to the FS algorithms, including ReliefF, JMI, DISR, ICAP, GiniI, MIM, CMIM, and CIFE, mRMR shows the computational efficiency at the third state. In addition, as the computational efficiencies of CFS, JMI, DISR, ICAP, GiniI, MIM, mRMR, CMIM, CIFE, and PmRMRE decrease with increasing data dimensionality, no obvious effects are shown by the ReliefF, FishS, and TraceR algorithms for all three datasets.

Based on the results shown in Figs. 3 and 4, we can conclude that PmRMRE is an ideal algorithm for constructing an EL

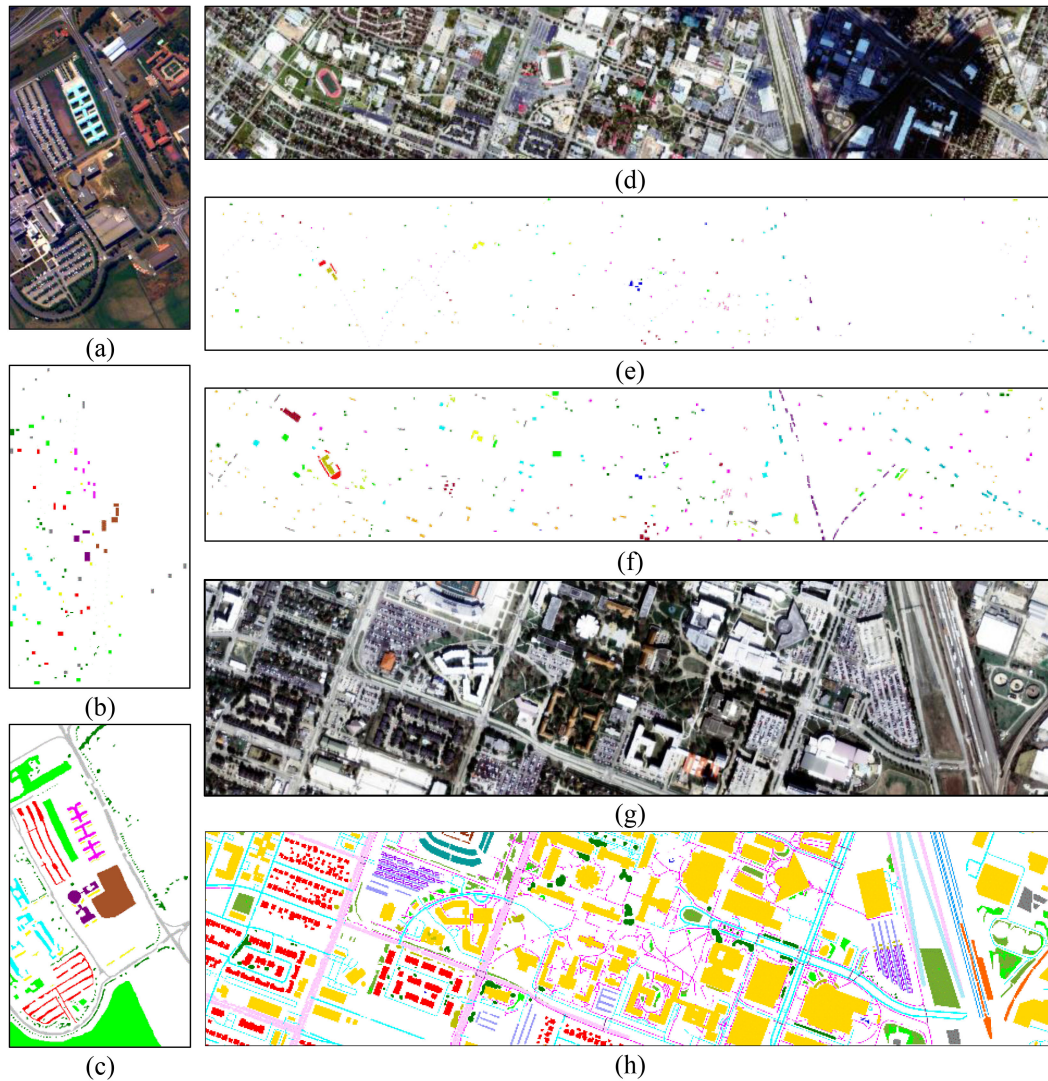


Fig. 2. Experimental datasets: (a) Pavia University; (d) GRSS-DFC2013 Houston; and (g) GRSS-DFC2018 Houston with the corresponding ground-truth maps (b), (c), (e), (f), and (h) used in this article.

system from the perspectives of both robustness and efficiency. However, there is high possibility that a robust FS algorithm (i.e., PmRMRE) returns highly similar and even exactly the same features in independent runs, which results in low classifier diversity and could limit the performance of the EL system. To avoid this issue, an incremental FS criterion was adopted. Considering the heterogeneous properties of the considered datasets on dimensionality and landscapes, the values of the incremental FS ranges ( $[\ell_{\min}, \ell_{\max}]$ ,  $\ell_{\max} \leq d$ ), as shown in Algorithm 2, need to be determined empirically.

According to the results in Fig. 3, we can see that RaF reached OA values higher than 68% by using ten raw spectral features (approximately one-tenth of the total number of spectral bands), and there were no obvious improvements when using more than 30 raw spectral features for Pavia University data [approximately one-third of the total number of spectral bands, see Fig. 3(a)]. Similarly, practically acceptable OA values (>65%) are reached by RaF when using only five selected raw spectral features (exactly one-ninth of the total numbers

of spectral bands), and obvious improvements are not available by using more than 20 features for the GRSS-DFC2018 data [see Fig. 3(k)]. Additionally, for the EMPs, EMPPR, and EMSER-MPsM features, approximately 8–10 features for the lower range and 25–30 features for the upper range can be noticed, as shown by Fig. 3(b)–(d) for the Pavia University data, Fig. 3(g)–(i) for the GRSS-DFC2013 Houston data, and Fig. 3(l)–(n) for the GRSS-DFC2018 Houston data. Based on the above results and for simple implementation, we roughly set the range to  $\ell_{\min} = d/10$  and  $\ell_{\max} = d/3$  for GPU-CatBF in the next experiments, where  $d$  is the feature dimensionality.

### B. Evaluation of the Proposed Method

Usually, the performance of a classifier is evaluated in terms of the classification accuracy and computational complexity. Hence, we show the OA values versus the ensemble sizes of the considered classifiers using various features from the considered

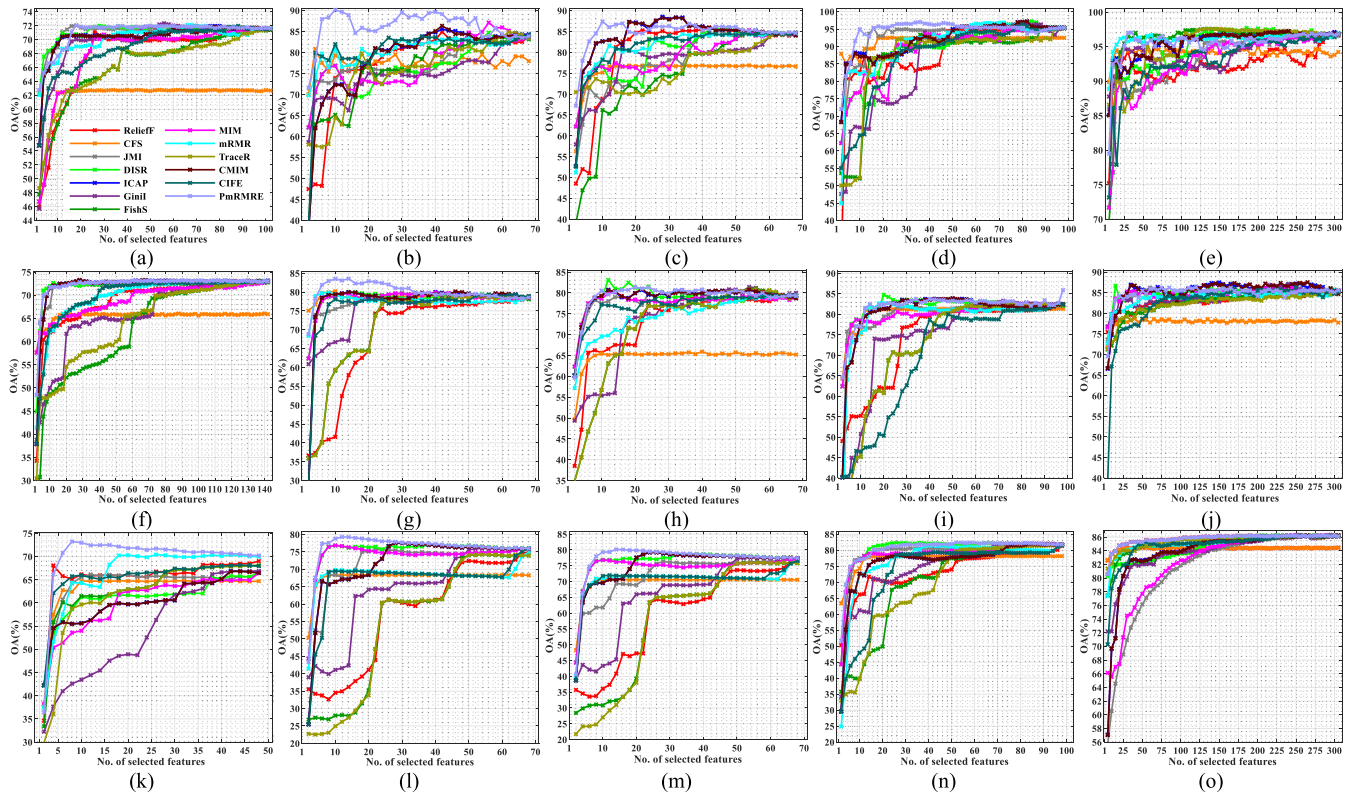


Fig. 3. RaF-based classification accuracy curves versus the number of selected features using considered feature selectors on various feature sets [(a), (f), (k): raw; (b), (g), (l): EMPs; (c), (h), (m): EMPPR; (d), (i), (n): EMSER-MPsM; (e), (j), (o): MRS-OO] from ROSIS University (row 1), GRSS-DFC2013 Houston (row 2), and GRSS-DFC2018 Houston (row 3). Each curve shows the mean OA from ten independent runs of experiments.

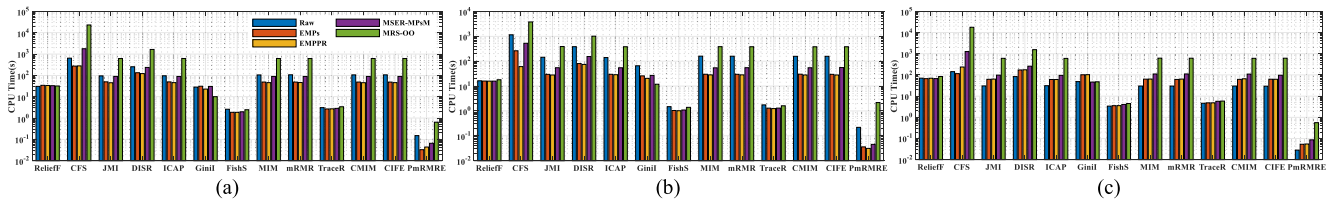


Fig. 4. Computational costs from considered feature selectors on various features from the (a) Pavia University, (b) GRSS-DFC2013 Houston, and (c) GRSS-DFC2018 Hyperspectral datasets.

datasets in Fig. 5, where the computational cost versus the ensemble size of the considered classifiers are shown in Fig. 6.

According to the results for the Pavia University data shown in the first row of Fig. 5, HistGBT (Cls1) and LightGBM (Cls11) showed the fastest convergence speed compared with the other classifiers by using the raw spectral features (see the learning curves in red and brown, respectively), while the worst results were attained by GPU-XGB-RaF (Cls10) (see the learning curve in olive green). Moreover, GPU-CatBF (Cls12) reaches the upper bound of the OA values when the ensemble size is greater than 150; see the learning curve in dark green. Compared with the results from CatBoost (Cls6) and GPU-CatBoost (Cls7), as shown by the learning curves in magenta and green, respectively, the learning curve from GPU-CatBF is always high. While the same results for GPU-CatBF are superior to those of CatBoost and GPU-CatBoost in terms of the classification accuracy and can also be found in Fig. 5(b)–(d) for the EMPs, EMPPR, and MRS-OO features, respectively, the highest OA values reached

by and GPU-CatBF using the EMPs and EMPPR features, as shown by the dark green learning curves in Fig. 5(b) and (c), particularly when the ensemble size is greater than 150. Additionally, CatBoost, GPU-CatBoost, and GPU-CatBF showed better capability of avoiding the well-known overfitting issue compared with AdaBoost (Cls4), GBDT (Cls5), and GPU-XGB-RaF (Cls10), as shown by the learning curves in spring green, blue, and cyan, respectively.

From the results shown in the second row of Fig. 5 for the GRSS-DFC2013 Houston data, it can be clearly seen that: 1) higher OA values are always available for GPU-CatBF compared with CatBoost and GPU-CatBoost; 2) GPU-CatBF shows the highest OA values in cases using the EMPs, EMPPR, and MRS-OO features; and 3) worse classification results caused by overfitting are obvious for XGBoost with RaF booster (Cls9) by using raw spectral, EMPs, EMPPR, and MRS-OO features, and for ExtraTrees (Cls3) using EMPPR features, as shown by learning curves in colors cyan and grey, respectively.



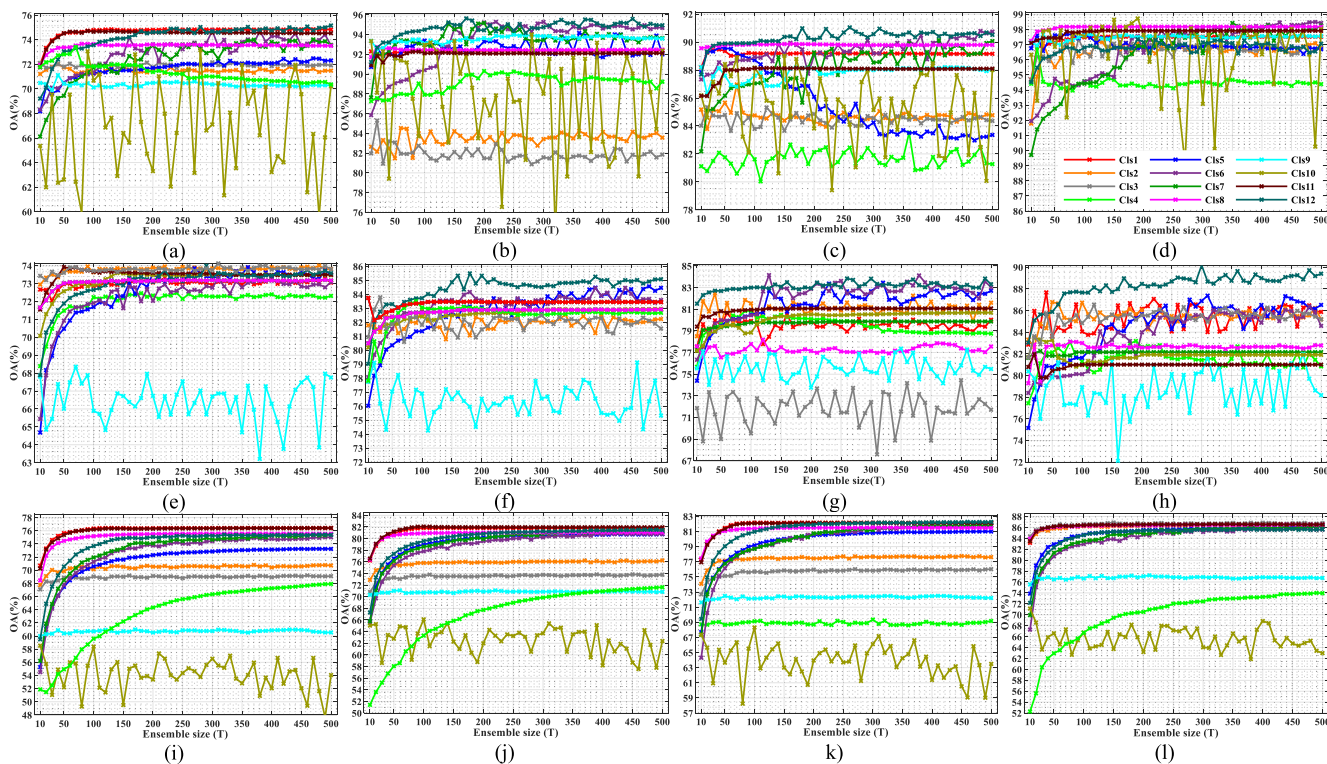


Fig. 5. Classification accuracy curves versus ensemble size of considered classifiers using various features [(a), (e), (i): raw; (b), (f), (j): EMPs; (c), (g), (k): EMPPR; (d), (h), (l): MRS-OO] from the Pavia University (row 1), GRSS-DFC2013 Houston (row 2), and GRSS-DFC2018 Houston (row 3) datasets.

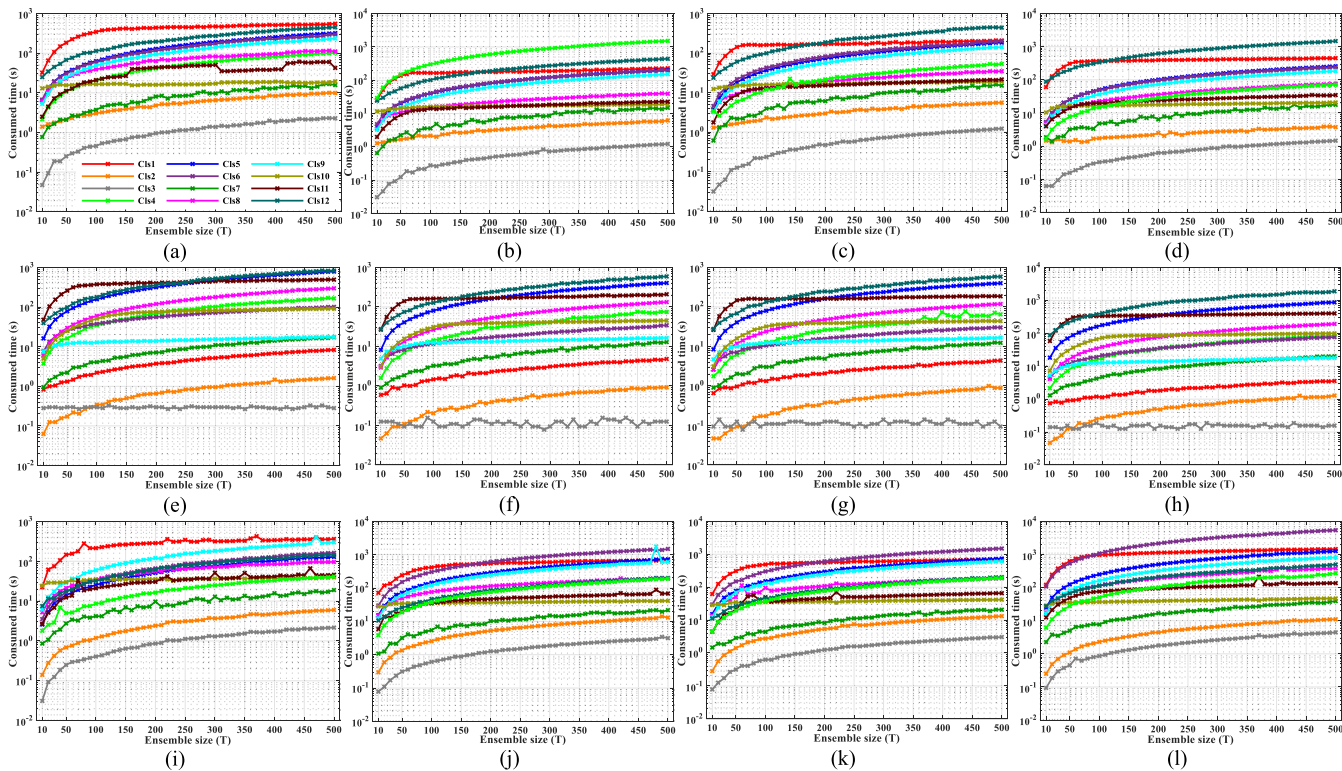


Fig. 6. Model training complexity curves versus ensemble size of considered classifiers using various features [(a), (e), (i): raw; (b), (f), (j): EMPs; (c), (g), (k): EMPPR; (d), (h), (l): MRS-OO] from the Pavia University (row 1), GRSS-DFC2013 Houston (row 2), and GRSS-DFC2018 Houston (row 3) datasets.

TABLE II  
OA, KAPPA, AND PREDICTION TIME VALUES FROM THE CONSIDERED CLASSIFIERS USING THE VARIOUS FEATURES OF THE PAVIA UNIVERSITY DATA (ENSEMBLE SIZE = 150)

No	Classifier	Raw		PC10		EMPs		EMPPR		SP-MPsM		EMSER-MPsM		MRS-OO	
		$d=103$		$d=10$		$d=70$		$d=70$		$d=100$		$d=100$		$d=310$	
		OA, K	Time (s)	OA, K	Time (s)	OA, K	Time (s)	OA, K	Time (s)	OA, K	Time (s)	OA, K	Time (s)	OA, K	Time (s)
Cls1	HistGBT	<b>74.79, 0.69</b>	<u>41.1875</u>	76.92, 0.71	<u>59.8125</u>	92.37, 0.90	<u>31.125</u>	89.19, 0.85	<u>29.4844</u>	87.15, 0.84	<u>28.75</u>	96.60, 0.96	<u>28.1094</u>	97.48, 0.97	35.4129
Cls2	RaF	71.68, 0.65	3.7969	76.79, 0.71	2.9219	85.33, 0.81	2.7969	85.36, 0.81	2.6094	86.15, 0.82	2.4531	95.63, 0.94	2.7813	96.72, 0.96	2.5
Cls3	ExtraTrees	71.94, 0.65	4.4375	78.64, 0.73	3.4219	81.29, 0.76	3.3281	84.13, 0.79	3.125	92.12, 0.90	2.8438	93.69, 0.92	3.1094	96.90, 0.96	2.6563
Cls4	AdaBoost	71.95, 0.64	24.7031	67.50, 0.60	25.0469	91.82, 0.89	23.9531	85.06, 0.80	23.375	90.12, 0.87	26.9219	<b>96.94, 0.96</b>	23.6719	94.67, 0.93	30.5
Cls5	GBDT	71.27, 0.65	5.375	73.74, 0.67	2.125	<b>93.42, 0.91</b>	4.9688	88.31, 0.84	4.875	90.69, 0.88	5.7188	95.57, 0.94	5.5625	97.18, 0.96	6.6094
Cls6	CatBoost	71.36, 0.65	0.6563	76.57, 0.71	0.4531	90.40, 0.87	0.5156	88.94, 0.85	0.7656	83.11, 0.79	0.5469	94.01, 0.92	0.6563	94.57, 0.93	0.75
Cls7	GPU-CatBoost	71.72, 0.65	0.7344	<b>78.66, 0.73</b>	0.9531	92.93, 0.91	0.6094	87.05, 0.83	0.5781	89.51, 0.87	0.7031	89.04, 0.86	0.6563	94.42, 0.93	0.8594
Cls8	XGB-CART	73.65, 0.68	13.8281	77.18, 0.71	12.6406	92.46, 0.90	9.8438	<b>89.85, 0.86</b>	8.2188	90.75, 0.88	8.6094	96.69, 0.96	10.1563	<b>98.18, 0.98</b>	11.7969
Cls9	XGB-RaF	70.12, 0.63	11.0625	73.68, 0.67	12.9063	93.10, 0.91	10.5313	87.60, 0.83	11.9063	88.90, 0.86	11.3281	94.03, 0.92	11.0156	97.63, 0.97	12.125
Cls10	GPU-XGB-RaF	70.68, 0.63	6.6563	69.28, 0.62	5.4688	92.97, 0.91	6.3594	88.17, 0.84	6.9375	<b>94.35, 0.93</b>	6.6563	94.73, 0.93	7.1094	97.54, 0.97	9.3438
Cls11	LightGBM	74.65, 0.69	21.4375	77.08, 0.71	21.3125	92.10, 0.90	18.1563	88.14, 0.84	18.6094	84.24, 0.80	17.9063	96.15, 0.95	18.625	97.88, 0.97	21.4531
Cls12	Proposed	73.37, 0.67	20.4375	~	~	92.75, 0.90	12.7188	89.03, 0.85	13.1719	91.11, 0.88	19.6406	95.95, 0.95	18.3125	95.75, 0.94	70.2031

TABLE III  
OA, KAPPA, AND PREDICTION TIME VALUES FROM THE CONSIDERED CLASSIFIERS USING THE VARIOUS FEATURES OF THE GRSS-DFC2013 HOUSTON DATA (ENSEMBLE SIZE = 150)

No	Classifier	Raw		PC10		EMPs		EMPPR		SP-MPsM		EMSER-MPsM		MRS-OO	
		$d=144$		$d=10$		$d=70$		$d=70$		$d=100$		$d=100$		$d=310$	
		OA, K	Time (s)	OA, K	Time (s)	OA, K	Time (s)	OA, K	Time (s)	OA, K	Time (s)	OA, K	Time (s)	OA, K	Time (s)
Cls1	HistGBT	<b>73.65, 0.72</b>	<u>203.516</u>	78.27, 0.77	<u>199.375</u>	83.47, 0.82	<u>172.875</u>	81.06, 0.80	<u>167.578</u>	80.40, 0.79	<u>167.078</u>	83.82, 0.83	<u>170.844</u>	81.00, 0.79	173.547
Cls2	RaF	72.97, 0.71	18.7031	78.36, 0.77	15.75	78.11, 0.76	15.4375	79.52, 0.78	14.75	82.34, 0.81	16.0313	82.73, 0.81	15.8594	86.94, 0.86	15.4844
Cls3	ExtraTrees	73.58, 0.72	20.5313	<b>79.76, 0.78</b>	17.8281	82.64, 0.81	17.2031	81.09, 0.80	16.3125	82.41, 0.81	17.2344	81.79, 0.80	17.2969	85.56, 0.84	15.6875
Cls4	AdaBoost	68.82, 0.66	99.4219	75.80, 0.74	104.609	84.25, 0.83	101.813	83.04, 0.82	101.344	81.58, 0.80	101.688	82.14, 0.81	101.797	86.43, 0.85	105.734
Cls5	GBDT	72.32, 0.70	52.7969	76.46, 0.75	22.9531	82.09, 0.81	44.2813	80.30, 0.79	44.2031	77.86, 0.76	50.0156	80.93, 0.79	49.2188	81.84, 0.80	57.7344
Cls6	CatBoost	72.30, 0.70	2.5469	77.92, 0.76	2.0625	82.71, 0.81	2.2656	81.97, 0.80	2.0781	81.96, 0.80	2.2344	83.80, 0.83	1.9063	84.72, 0.83	3.3281
Cls7	GPU-CatBoost	72.01, 0.70	2.8438	78.57, 0.77	2.6094	84.48, 0.83	2.6094	83.05, 0.82	2.625	82.52, 0.81	2.75	82.31, 0.81	2.7813	83.64, 0.82	3.1406
Cls8	XGB-CART	73.17, 0.71	66.4063	77.73, 0.76	56.9688	82.81, 0.81	50.6406	79.77, 0.78	47.1406	82.09, 0.81	49.125	84.59, 0.83	53.1875	82.16, 0.81	55.7656
Cls9	XGB-RaF	70.04, 0.68	51.0781	74.89, 0.73	50.4531	79.63, 0.78	47.3281	77.79, 0.76	43.875	76.95, 0.75	47.5469	78.56, 0.77	47.72	83.00, 0.82	55.875
Cls10	GPU-XGB-RaF	66.09, 0.64	31.0469	70.56, 0.68	24.3281	77.67, 0.76	27.0313	75.24, 0.73	27.5469	75.43, 0.73	27.5469	77.82, 0.76	28.75	79.09, 0.77	38.6406
Cls11	LightGBM	73.61, 0.72	101.875	78.48, 0.77	104.594	82.78, 0.81	94.2344	82.16, 0.81	93.125	79.99, 0.78	91.4844	83.77, 0.82	93.7813	83.73, 0.82	88.6719
Cls12	Proposed	73.37, 0.71	93.875	~	~	<b>85.09, 0.84</b>	46.6094	<b>83.16, 0.82</b>	46.3594	<b>85.05, 0.84</b>	65.0694	<b>85.15, 0.84</b>	65.1875	<b>88.60, 0.88</b>	<u>221.484</u>

Looking at the graphs from the last row of Fig. 5, the best results with the highest OA values are reached by HistGBT (Cls1), LightGBM (Cls11), and XGBoost with CART booster (Cls8) on raw spectral, EMPs, EMPPR, and MRS-OO features, while better than results from RaF (Cls2), ExtraTrees (Cls3), AdaBoost (Cls4), GBDT (Cls5), CatBoost (Cls6), GPU-CatBoost (Cls7), XGBoost-RaF (Cls9), GPU-accelerated XGBoost-RaF (Cls10) are obtained by GPU-CatBF (Cls12) on raw spectra, EMPs, and EMPPR features. Additionally, the OA of GPU-CatBF is maximized only when the ensemble size is greater than 150.

According to the computational complexity cost results shown in Fig. 6, the following results can be observed. First and foremost, RaF (Cls2, orange curves) and ExtraTrees (Cls3, grey curves) are the most efficient algorithms in all cases compared with the other classifiers. Moreover, GPU-CatBoost (Cls7) is always at least ten times faster than the CPU-based implementation of CatBoost (Cls6), as expected, as depicted by the cost curves in green and magenta, respectively. Furthermore, greater computational efficiency than HistGBT (Cls1) is reachable for GPU-CatBF (Cls12) on the Pavia University and GRSS-DFC2018 Houston datasets, particularly if the ensemble size is approximately less than 150. GPU-CatBF is more efficient than LightGBM (Cls11) on the GRSS-2013 Houston data only when the ensemble size is less than 100. Finally, HistGBT (Cls1, red curves) is at least 100 times faster than the original LightGBM version when large amounts of training data are available (see the graphs in the second row of Fig. 6), which is in accordance with the assumption in [73].

Summarizing the above results drawn from Figs. 5 and 6, we empirically set the ensemble size of the proposed GPU-CatBF algorithm to 150 for both accurate and efficient classification.

### C. Classification Results Comparison

To further compare the proposed classification approach for land cover mapping using hyperspectral images, Tables II–IV show the classification OA,  $\kappa$ , and prediction time cost in seconds for the considered classifiers using different types of features, and Figs. 7 and 8 present the final best classification maps with the OA values highlighted in boldface from Tables II–IV. For a fair comparison, the ensemble size for all the classifiers is set to the same value of 150, which is the recommended value of the ensemble size for CatBoost, CPU-CatBoost, and the proposed GPU-CatBF algorithm.

Again, it can be clearly seen that the HistGBT classifier reaches the highest classification values for raw spectral features from all three test images; see the highlighted numbers in bold from the third column of Tables IV–VI. However, HistGBT is slower than all the other classifiers in the prediction phase, specifically using the raw spectral, first ten principal components (PC10), EMPs, EMPPR, SP-MPsM, and EMSER-MPsM features from the Pavia University and GRSS-DFC2018 Houston test images and using all features from the GRSS-DFC2013 test image; see the underlined numbers from the first row of Tables II–IV.

TABLE IV  
OA, KAPPA, AND PREDICTION TIME VALUES FROM THE CONSIDERED CLASSIFIERS USING THE VARIOUS FEATURES OF THE GRSS-DFC2018 HOUSTON DATA (ENSEMBLE SIZE = 150)

No	Classifier	Raw		PC10		EMPs		EMPPR		SP-MPsM		EMSER-MPsM		MRS-OO	
		$d=48$		$d=10$		$d=70$		$d=70$		$d=100$		$d=100$		$d=310$	
		OA, K	Time (s)	OA, K	Time (s)	OA, K	Time (s)	OA, K	Time (s)	OA, K	Time (s)	OA, K	Time (s)	OA, K	Time (s)
Cls1	HistGBT	<b>76.33, 0.69</b>	1960.84	78.45, 0.72	1570.53	81.81, 0.76	1927.61	<b>82.17, 0.77</b>	1648.80	86.13, 0.82	1627.64	85.84, 0.82	1957.13	86.26, 0.82	1972.36
Cls2	RaF	70.41, 0.63	141.875	76.89, 0.70	117.063	76.59, 0.69	130.781	77.27, 0.71	130.875	83.10, 0.78	127.734	81.81, 0.76	129.563	86.21, 0.82	121.094
Cls3	ExtraTrees	68.92, 0.61	156.516	77.30, 0.71	130.859	73.52, 0.66	149.656	75.62, 0.69	144.484	82.37, 0.77	135.922	80.88, 0.75	138.641	<b>86.52, 0.82</b>	122.531
Cls4	AdaBoost	63.32, 0.54	987	64.86, 0.56	997.719	65.91, 0.56	993.859	70.64, 0.62	998.078	76.29, 0.69	993.641	70.55, 0.61	992.969	85.70, 0.81	979.969
Cls5	GBDT	70.58, 0.62	340.813	74.50, 0.67	227.766	78.96, 0.73	387.938	79.16, 0.73	386.453	84.50, 0.80	412.094	83.71, 0.79	409.25	84.66, 0.80	450.453
Cls6	CatBoost	71.28, 0.64	14.5	75.48, 0.69	12.6563	77.82, 0.72	15.2813	78.73, 0.73	14.9219	81.35, 0.76	16.2344	81.24, 0.76	15.7656	82.95, 0.78	21.5469
Cls7	GPU-CatBoost	71.94, 0.64	13.9844	76.05, 0.69	12.4688	78.59, 0.73	14.7969	78.71, 0.73	14.5625	82.20, 0.77	15.9531	81.53, 0.76	16.1406	83.50, 0.79	21.0781
Cls8	XGB-CART	75.17, 0.68	764.438	77.46, 0.71	758.359	80.91, 0.75	729.906	81.39, 0.76	716.078	85.31, 0.81	692.109	85.17, 0.81	717.359	86.30, 0.82	804.672
Cls9	XGB-RaF	60.70, 0.52	882.203	67.85, 0.59	975.641	70.71, 0.63	906.109	72.14, 0.65	913.671	76.78, 0.71	957.156	72.83, 0.66	930.531	76.65, 0.69	1058.34
Cls10	GPU-XGB-RaF	58.39, 0.44	347.297	62.86, 0.50	324.75	66.11, 0.55	361.281	68.32, 0.58	360.609	70.59, 0.61	378.078	67.06, 0.55	382.156	66.74, 0.54	486.25
Cls11	LightGBM	76.20, 0.69	1166.53	<b>78.47, 0.72</b>	1163.23	<b>82.06, 0.77</b>	1115.92	82.08, 0.77	1136.23	<b>86.16, 0.92</b>	1103.64	<b>86.12, 0.82</b>	1115.44	86.38, 0.82	1162.75
Cls12	Proposed	75.44, 0.69	213.766	~	~	81.16, 0.76	299.641	81.80, 0.77	300.156	84.37, 0.80	425.016	83.17, 0.78	425.563	85.70, 0.81	1214.61

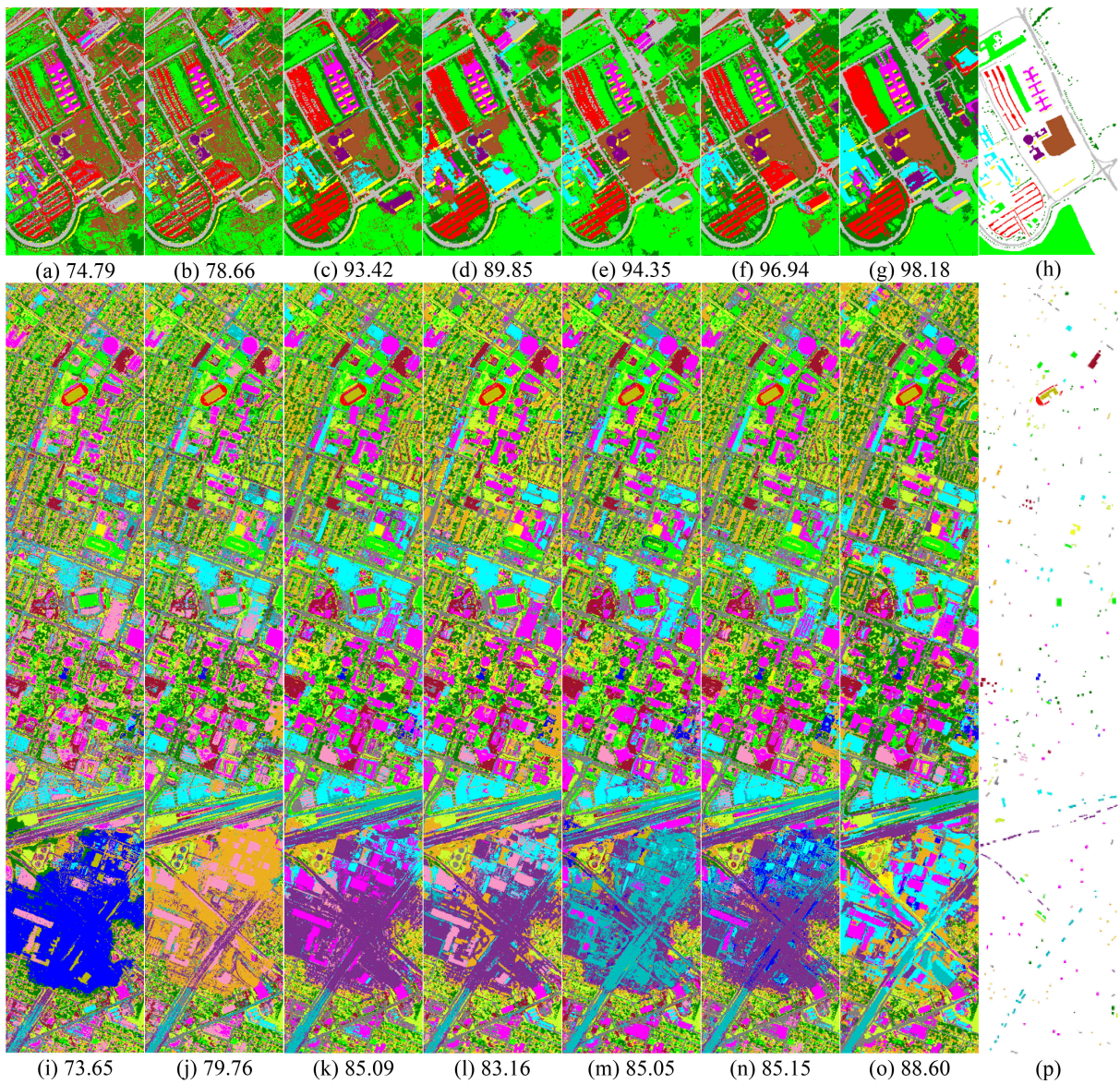


Fig. 7. Classification maps with OA values corresponding to the numbers in bold from (a)–(g) Table II and (i)–(o) Table III and ground-truth maps for the (h) Pavia University (column 1) and (p) GRSS-DFC2013 (column 2) datasets.

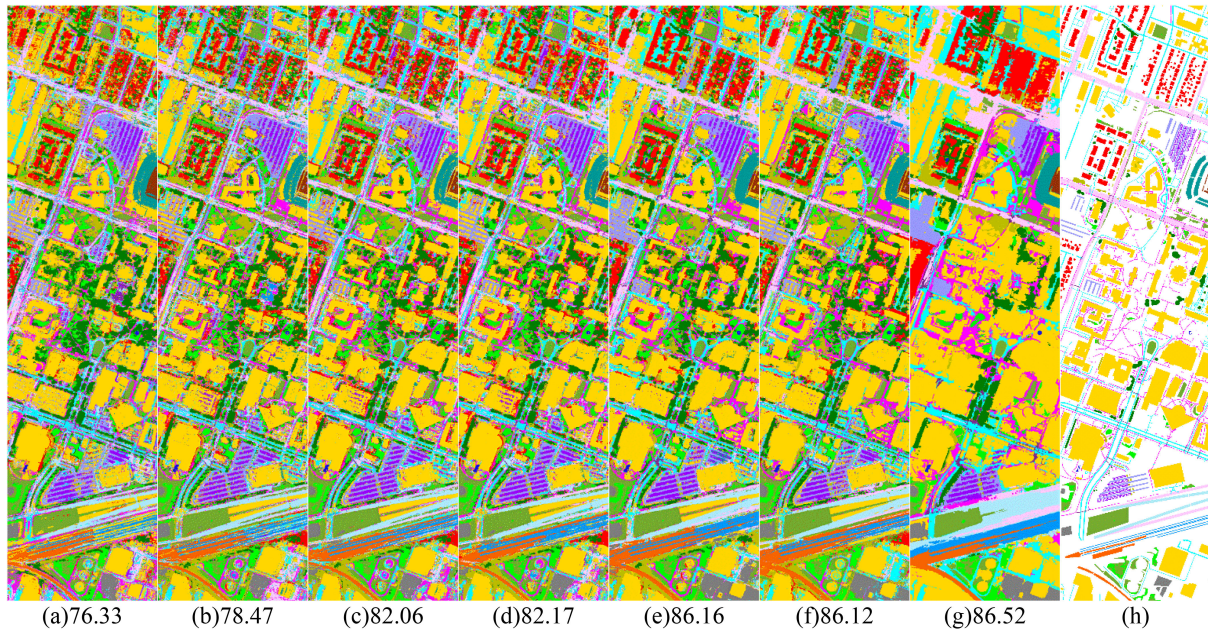


Fig. 8. Classification maps with OA values correspondent to the numbers in bold from (a)–(g) Table IV and the (h) ground-truth map for the GRSS-DFC2018 Houston data.

GPU-CatBF achieves higher OA values than CatBoost and GPU-CatBoost, but prediction is slower. This finding is universally true for all cases of using various features from the considered test images and is in accordance with the previous results shown in Figs. 5 and 6. In addition, compared with the results from the other classifiers, GPU-CatBF is capable of reaching compatibility and achieves an even better classification accuracy. For instance, GPU-CatBF reached the highest OA values by using EMPs, EMPPR, SP-MPsM, EMSER-MPsM, and MRS-OO features from the GRSS-DFC2013 Houston data; see the numbers in bold in the last row of Table III. If we compare the results from the Pavia University data, although the best results are not obtained by GPU-CatBF, OA values higher than those from other classifiers are present. For example, HistGBT and LightGBM obtained OA values of 92.37% and 92.10%, respectively, when using EMPs features, while an OA value of 92.75% was reached by GPU-CatBF in a shorter prediction time. From the results from the GRSS-DFC2018 Houston test data, although the highest OA values are reached by either HistGBT or LightGBM in most cases, OA values higher than those from GBDT and XGB-CART can be observed for GPU-CatBF using the raw spectral, EMPs, EMPPR, and MRS-OO features.

In our previous works in [25], [26], [71], and [72], the performances of EMSER-MPsM, SP-MPsM, and MRS-OO were separately investigated by comparison with the MPs, EMPPR, EMPs, and EMPPR features. Hence, it is worth comprehensively comparing their performances here. First, according to the results shown in Tables II–IV and Figs. 6 and 7, it is clear that better classification results with higher OA values can be obtained by the SP-MPsM and EMSER-MPsM features than with the EMPs and EMPPR features. For example, while all the considered classifiers reached OA values between 89.04% and 96.94% by using the EMSER-MPsM features from the Pavia University

data, the OA value ranges shown by the considered classifiers are between 85.33% and 93.42% and between 84.13% and 89.85 when using the EMPs and EMPPR features, respectively. Furthermore, when we compare the classification results from the SP-MPsM, EMSER-MPsM, and MRS-OO features, generally better classification results are obtained by EMSER-MPsM in contrast with the results from SP-MPsM features for Pavia University and GRSS-DFC2013 Houston test images, while the best results are obtained from MRS-OO features for all the test images. For instance, see that the area in the lower part of the GRSS-DFC2013 Houston image, which is covered by dense cloud shadows, is more precisely classified by using the MRS-OO features.

## VI. CONCLUSION

In this article, the GPU-CatBoost algorithm for hyperspectral image classification was introduced and comparatively studied in terms of the classification accuracy and computational efficiency using diverse features. To further boost the classification performance by considering its highly accelerated advantage with respect to the CPU-based implementation of CatBoost, an incremental subspace FS-based ensemble version, GPU-CatBF, is proposed. To evaluate the performance of the proposed approach, 11 popular DT-based EL algorithms, namely, the HistGBT, RaF, ExtraTrees, AdaBoost, GBDT, CatBoost, GPU-CatBoost, XGB-CART, XGB-RaF, GPU-XGB-RaF, and LightGBM algorithms, are selected in the experiments. Moreover, 11 popular FS algorithms, namely, the ReliefF, CFS, JMI, DISR, ICAP, GiniI, FishS, MIM, TraceR, CMIM, and CIFE methods, are selected to evaluate the performance of mRMR and PmRMRE. According to the experimental results from three

widely used hyperspectral benchmarks, the concluded results are as follows.

- 1) Compared with existing popular FS algorithms, the superior properties of mRMR and PmRMRE for highly discriminative subspace FS from hyperspectral images categorized with diverse feature sets are clear, while the best results are shown by PmRMRE in the context of both the robustness and computational efficiency.
- 2) Compared with popular DT-based EL algorithms and the CPU-based implementation of CatBoost, GPU-CatBoost is also an advanced EL algorithm for hyperspectral image classification using various features.
- 3) While further improved performance of the proposed GPU-CatBF is clear compared with CatBoost and GPU-CatBoost, similar and even better classification accuracy results are reachable for GPU-CatBF in some cases.

Although GPU-CatBF outperforms CatBoost and GPU-CatBoost in terms of the classification performance, the computational cost from a larger value set of the ensemble size is also clear. Additionally, as an advanced FS algorithm, the computational efficiency of PmRMRE could be further enhanced in its GPU-based implementation. Therefore, we will focus on the self-adaptive selection of the ensemble size of GPU-CatBoost, GPU acceleration of PmRMRE, and mixed-precision technique in future work.

#### ACKNOWLEDGMENT

The authors would like to thank Prof. P. Gamba, the IEEE GRSS Image Analysis and DFCT, and the NCALM and the Hyperspectral Image Analysis Laboratory, University of Houston, for freely providing the Pavia University, GRSS-DFC2013 Houston, and GRSS-DFC2018 Houston hyperspectral datasets used in this article.

#### REFERENCES

- [1] T. Tolessa, F. Senbeta, and M. Kidane, "The impact of land use/land cover change on ecosystem services in the central highlands of Ethiopia," *Ecosyst. Serv.*, vol. 23, pp. 47–54, 2017.
- [2] R. A. Houghton and A. A. Nassikas, "Global and regional fluxes of carbon from land use and land cover change 1850–2015," *Global Biogeochem. Cycles*, vol. 31, no. 3, pp. 456–472, 2017.
- [3] D. Lu and Q. Weng, "A survey of image classification methods and techniques for improving classification performance," *Int. J. Remote Sens.*, vol. 28, no. 5, pp. 823–870, 2007.
- [4] U. Maulik and D. Chakraborty, "Remote sensing image classification: A survey of support-vector-machine-based advanced techniques," *IEEE Geosci. Remote Sens. Mag.*, vol. 5, no. 1, pp. 33–52, Mar. 2017.
- [5] X. X. Zhu *et al.*, "Deep learning in remote sensing: A comprehensive review and list of resources," *IEEE Geosci. Remote Sens. Mag.*, vol. 5, no. 4, pp. 8–36, Dec. 2017.
- [6] M. Belgiu and L. Drăguț, "Random forest in remote sensing: A review of applications and future directions," *ISPRS J. Photogramm. Remote Sens.*, vol. 114, pp. 24–31, 2016.
- [7] B. M. Shahshahani and D. A. Landgrebe, "The effect of unlabeled samples in reducing the small sample size problem and mitigating the Hughes phenomenon," *IEEE Trans. Geosci. Remote Sens.*, vol. 32, no. 5, pp. 1087–1095, Sep. 1994.
- [8] G. Camps-Valls, D. Tuia, L. Bruzzone, and J. A. Benediktsson, "Advances in hyperspectral image classification: Earth monitoring with statistical learning methods," *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 45–54, Jan. 2014.
- [9] G. M. Foody, A. Mathur, C. Sanchez-Hernandez, and D. S. Boyd, "Training set size requirements for the classification of a specific class," *Remote Sens. Environ.*, vol. 104, no. 1, pp. 1–14, 2006.
- [10] G. M. Foody and A. Mathur, "The use of small training sets containing mixed pixels for accurate hard image classification: Training on mixed spectral responses for classification by a SVM," *Remote Sens. Environ.*, vol. 103, no. 2, pp. 179–189, 2006.
- [11] C. A. Lee, S. D. Gasster, A. Plaza, C. I. Chang, and B. Huang, "Recent developments in high performance computing for remote sensing: A review," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 4, no. 3, pp. 508–527, Sep. 2011.
- [12] J. F. Mas and J. J. Flores, "The application of artificial neural networks to the analysis of remotely sensed data," *Int. J. Remote Sens.*, vol. 29, no. 3, pp. 617–663, 2008.
- [13] A. Samat, P. Du, S. Liu, J. Li, and L. Cheng, "E2LMs: Ensemble extreme learning machines for hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 4, pp. 1060–1069, Apr. 2014.
- [14] V. F. Rodriguez-Galiano, B. Ghimire, J. Rogan, M. Chica-Olmo, and J. P. Rigol-Sanchez, "An assessment of the effectiveness of a random forest classifier for land-cover classification," *ISPRS J. Photogramm. Remote Sens.*, vol. 67, pp. 93–104, 2012.
- [15] P. Du, A. Samat, B. Waske, S. Liu, and Z. Li, "Random forest and rotation forest for fully polarized SAR image classification using polarimetric and spatial features," *ISPRS J. Photogramm. Remote Sens.*, vol. 105, pp. 38–53, 2015.
- [16] D. Hong *et al.*, "More diverse means better: Multimodal deep learning meets remote-sensing imagery classification," *IEEE Trans. Geosci. Remote Sens.*, to be published, doi: [10.1109/TGRS.2020.3016820](https://doi.org/10.1109/TGRS.2020.3016820).
- [17] D. Hong, J. Yao, D. Meng, Z. Xu, and J. Chanussot, "Multimodal GANs: Toward crossmodal hyperspectral-multispectral image segmentation," *IEEE Trans. Geosci. Remote Sens.*, to be published, doi: [10.1109/TGRS.2020.3020823](https://doi.org/10.1109/TGRS.2020.3020823).
- [18] P. Du *et al.*, "Advances of four machine learning methods for spatial data handling: A review," *J. Geovis. Spatial Anal.*, vol. 4, 2020, Art. no. 13.
- [19] G. Camps-Valls, L. Gómez-Chova, J. Muñoz-Marí, J. L. Rojo-Álvarez, and M. Martínez-Ramón, "Kernel-based framework for multitemporal and multisource remote sensing data classification and change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 6, pp. 1822–1835, Jun. 2008.
- [20] J. C. W. Chan, C. Huang, and R. DeFries, "Enhanced algorithm performance for land cover classification from remotely sensed data using bagging and boosting," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 3, pp. 693–695, Mar. 2001.
- [21] J. C. W. Chan and D. Paelinckx, "Evaluation of random forest and Adaboost tree-based ensemble classification and spectral band selection for ecotope mapping using airborne hyperspectral imagery," *Remote Sens. Environ.*, vol. 112, no. 6, pp. 2999–3011, 2008.
- [22] J. Xia, P. Du, X. He, and J. Chanussot, "Hyperspectral remote sensing image classification based on rotation forest," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 1, pp. 239–243, Jan. 2014.
- [23] K. Millard and M. Richardson, "Wetland mapping with LiDAR derivatives, SAR polarimetric decompositions, and LiDAR–SAR fusion using a random forest classifier," *Can. J. Remote Sens.*, vol. 39, no. 4, pp. 290–307, 2013.
- [24] O. S. Ahmed, S. E. Franklin, M. A. Wulder, and J. C. White, "Characterizing stand-level forest canopy cover and height using Landsat time series, samples of airborne LiDAR, and the random forest algorithm," *ISPRS J. Photogramm. Remote Sens.*, vol. 101, pp. 89–101, 2015.
- [25] A. Samat, C. Persello, S. Liu, E. Li, Z. Miao, and J. Abuduwaili, "Classification of VHR multispectral images using extratrees and maximally stable extremal region-guided morphological profile," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 9, pp. 3179–3195, Sep. 2018.
- [26] A. Samat, E. Li, W. Wang, S. Liu, C. Lin, and J. Abuduwaili, "Meta-XGBoost for hyperspectral image classification using extended MSER-guided morphological profiles," *Remote Sens.*, vol. 12, no. 12, 2020, Art. no. 1973.
- [27] Y. Boualleg, M. Farah, and I. R. Farah, "Remote sensing scene classification using convolutional features and deep forest classifier," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 12, pp. 1944–1948, Dec. 2019.
- [28] X. Cao, L. Wen, Y. Ge, J. Zhao, and L. Jiao, "Rotation-based deep forest for hyperspectral imagery classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 7, pp. 1105–1109, Jul. 2019.

- [29] J. Ye, J. H. Chow, J. Chen, and Z. Zheng, "Stochastic gradient boosted distributed decision trees," in *Proc. 18th ACM Conf. Inf. Knowl. Manage.*, 2009, pp. 2061–2064.
- [30] T. Chen, T. He, M. Benesty, V. Khotilovich, and Y. Tang, "Xgboost: Extreme gradient boosting," R Package Version 0.4-2., vol. 1, no. 4, Aug. 2015.
- [31] G. Ke *et al.*, "LightGBM: A highly efficient gradient boosting decision tree," in *Proc. Neural Inf. Process. Syst. Conf.*, 2017, pp. 3146–3154.
- [32] S. Aiello, E. Eckstrand, A. Fu, M. Landry, and P. Aboyou, "Machine learning with R and H<sub>2</sub>O," in *H<sub>2</sub>O Booklet*. Mountain View, CA, USA: H2O.ai, 2016, p. 550.
- [33] A. V. Dorogush, V. Ershov, and A. Gulin, "CatBoost: Gradient boosting with categorical features support," 2018, *arXiv:1810.11363*.
- [34] A. Samat, E. Li, P. Du, S. Liu, Z. Miao, and W. Zhang, "CatBoost for RS image classification with pseudo label support from neighbor patches based clustering," *IEEE Geosci. Remote Sens. Lett.*, to be published, doi: [10.1109/LGRS.2020.3038771](https://doi.org/10.1109/LGRS.2020.3038771).
- [35] Z. Wu *et al.*, "GPU parallel implementation of spatially adaptive hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 4, pp. 1131–1143, Apr. 2017.
- [36] C. Gonzalez, S. Sánchez, A. Paz, J. Resano, D. Mozos, and A. Plaza, "Use of FPGA or GPU-based architectures for remotely sensed hyperspectral image processing," *Integration*, vol. 46, no. 2, pp. 89–103, 2013.
- [37] A. Yusuf and S. Alawneh, "A survey of GPU implementations for hyperspectral image classification in remote sensing," *Can. J. Remote Sens.*, vol. 44, no. 5, pp. 532–550, 2018.
- [38] R. Mitchell, A. Adinets, T. Rao, and E. Frank, "Xgboost: Scalable GPU accelerated learning," 2018, *arXiv:1806.11248*.
- [39] A. Plaza, Q. Du, Y. L. Chang, and R. L. King, "High performance computing for hyperspectral remote sensing," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 4, no. 3, pp. 528–544, Sep. 2011.
- [40] K. Jansson, H. Sundell, and H. Boström, "gpuRF and gpuERT: Efficient and scalable GPU algorithms for decision tree ensembles," in *Proc. IEEE Int. Parallel Distrib. Process. Symp. Workshops*, 2014, pp. 1612–1621.
- [41] P. Du, J. Xia, W. Zhang, K. Tan, Y. Liu, and S. Liu, "Multiple classifier system for remote sensing image classification: A review," *Sensors*, vol. 12, no. 4, pp. 4764–4792, 2012.
- [42] A. Samat, P. Du, M. H. A. Baig, S. Chakravarty, and L. Cheng, "Ensemble learning with multiple classifiers and polarimetric features for polarized SAR image classification," *Photogramm. Eng. Remote Sens.*, vol. 80, no. 3, pp. 239–251, 2014.
- [43] J. J. Rodriguez, L. I. Kuncheva, and C. J. Alonso, "Rotation forest: A new classifier ensemble method," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 10, pp. 1619–1630, Oct. 2006.
- [44] L. I. Kuncheva, J. J. Rodríguez, C. O. Plumptre, D. E. Linden, and S. J. Johnston, "Random subspace ensembles for fMRI classification," *IEEE Trans. Med. Imag.*, vol. 29, no. 2, pp. 531–542, Feb. 2010.
- [45] J. Xia, M. D. Mura, J. Chanussot, P. Du, and X. He, "Random subspace ensembles for hyperspectral image classification with extended morphological attribute profiles," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 9, pp. 4768–4786, Sep. 2015.
- [46] D. Lungu, S. Prasad, M. M. Crawford, and O. Ersoy, "Manifold-learning-based feature extraction for classification of hyperspectral data: A review of advances in manifold learning," *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 55–66, Jan. 2014.
- [47] W. Sun and Q. Du, "Hyperspectral band selection: A review," *IEEE Geosci. Remote Sens. Mag.*, vol. 7, no. 2, pp. 118–139, Jun. 2019.
- [48] Y. Y. Tang, H. Yuan, and L. Li, "Manifold-based sparse representation for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 12, pp. 7606–7618, Dec. 2014.
- [49] K. Tan, E. Li, Q. Du, and P. Du, "Hyperspectral image classification using band selection and morphological profiles," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 1, pp. 40–48, Jan. 2014.
- [50] R. Sheikhpour, M. A. Sarram, S. Gharaghani, and M. A. Z. Chahooki, "A survey on semi-supervised feature selection methods," *Pattern Recognit.*, vol. 64, pp. 141–158, 2017.
- [51] Y. Wang *et al.*, "Self-supervised feature learning with CRF embedding for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 5, pp. 2628–2642, May 2019.
- [52] S. Ramírez-Gallego *et al.*, "Fast-mRMR: Fast minimum redundancy maximum relevance algorithm for high-dimensional big data," *Int. J. Intell. Syst.*, vol. 32, no. 2, pp. 134–152, 2017.
- [53] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 8, pp. 1226–1238, Aug. 2005.
- [54] C. Ding and H. Peng, "Minimum redundancy feature selection from microarray gene expression data," *J. Bioinf. Comput. Biol.*, vol. 3, no. 2, pp. 185–205, 2005.
- [55] Z. Li, X. Zhou, Z. Dai, and X. Zou, "Classification of G-protein coupled receptors based on support vector machine with maximum relevance minimum redundancy and genetic algorithm," *BMC Bioinform.*, vol. 11, no. 1, 2010, Art. no. 325.
- [56] M. Radovic, M. Ghalwash, N. Filipovic, and Z. Obradovic, "Minimum redundancy maximum relevance feature selection approach for temporal gene expression data," *BMC Bioinform.*, vol. 18, no. 1, pp. 1–14, 2017.
- [57] M. I. Hejazi and X. Cai, "Input variable selection for water resources systems using a modified minimum redundancy maximum relevance (mMRMR) algorithm," *Adv. Water Resour.*, vol. 32, no. 4, pp. 582–593, 2009.
- [58] H. Rabi, M. I. Saripan, S. Mashohor, and M. H. Marhaban, "3D facial expression recognition using maximum relevance minimum redundancy geometrical features," *EURASIP J. Adv. Signal Process.*, vol. 2012, no. 1, 2012, Art. no. 213.
- [59] M. Toğaçar, B. Ergen, and Z. Cömert, "Detection of lung cancer on chest CT images using minimum redundancy maximum relevance feature selection method with convolutional neural networks," *Biocybern. Biomed. Eng.*, vol. 40, no. 1, pp. 23–39, 2020.
- [60] Y. Fan, X. Li, and P. Zhang, "Real-time static voltage stability assessment in large-scale power systems based on maximum-relevance minimum-redundancy ensemble approach," *IEEE Access*, vol. 5, pp. 27281–27291, 2017.
- [61] D. Bratasanu, I. Nedelcu, and M. Datu, "Interactive spectral band discovery for exploratory visual analysis of satellite images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 1, pp. 207–224, Feb. 2012.
- [62] M. Kamandar and H. Ghassemian, "Maximum relevance, minimum redundancy band selection for hyperspectral images," in *Proc. 19th Iranian Conf. Elect. Eng.*, 2011, pp. 1–5.
- [63] Y. Liu *et al.*, "Maximum relevance, minimum redundancy band selection based on neighborhood rough set for hyperspectral data classification," *Meas. Sci. Technol.*, vol. 27, no. 12, 2016, Art. no. 125501.
- [64] Y. Liu, M. Li, S. Wang, R. Wang, and W. Jiang, "Hyperspectral waveband selection algorithm based on weighted maximum relevance minimum redundancy and its stability analysis," *Meas. Sci. Technol.*, vol. 31, no. 8, 2020, Art. no. 085501.
- [65] S. Ramírez-Gallego *et al.*, "Fast-mRMR: Fast minimum redundancy maximum relevance algorithm for high-dimensional big data," *Int. J. Intell. Syst.*, vol. 32, no. 2, pp. 134–152, 2017.
- [66] R. M. N. Marone, F. Camara, and S. Ndiaye, "A parallelized spark based version of mRMR," in *Proc. Int. Conf. Innov. Interdisciplinary Solutions Underserved Areas*, 2018, pp. 187–198.
- [67] N. De Jay, S. Papillon-Cavanagh, C. Olsen, N. El-Hachem, G. Bontempi, and B. Haibe-Kains, "mRMR: An R package for parallelized mRMR ensemble feature selection," *Bioinformatics*, vol. 29, no. 18, pp. 2365–2368, 2013.
- [68] C. Zhang and Y. Ma, Eds., *Ensemble Machine Learning: Methods and Applications*. Berlin, Germany: Springer, 2012.
- [69] J. A. Benediktsson, J. A. Palmason, and J. R. Sveinsson, "Classification of hyperspectral data from urban areas based on extended morphological profiles," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 480–491, Mar. 2005.
- [70] W. Liao *et al.*, "Taking optimal advantage of fine spatial resolution: Promoting partial image reconstruction for the morphological analysis of very-high-resolution images," *IEEE Geosci. Remote Sens. Mag.*, vol. 5, no. 2, pp. 8–28, Jun. 2017.
- [71] A. Samat, S. Liu, C. Persello, E. Li, Z. Miao, and J. Abuduwaili, "Evaluation of ForestPA for VHR RS image classification using spectral and superpixel-guided morphological profiles," *Eur. J. Remote Sens.*, vol. 52, no. 1, pp. 107–121, 2019.
- [72] A. Samat *et al.*, "Direct, ECOC, ND and END frameworks—Which one is the best? An empirical study of Sentinel-2A MSILIC image classification for arid-land vegetation mapping in the Ili river delta, Kazakhstan," *Remote Sens.*, vol. 11, no. 16, 2019, Art. no. 1953.
- [73] F. Pedregosa *et al.*, "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2011.
- [74] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.
- [75] P. Geurts, D. Ernst, and L. Wehenkel, "Extremely randomized trees," *Mach. Learn.*, vol. 63, no. 1, pp. 3–42, 2006.

- [76] I. Mukherjee and R. E. Schapire, "A theory of multiclass boosting," *J Mach. Learn. Res.*, vol. 14, pp. 437–497, 2013.
- [77] M. Robnik-Šikonja and I. Kononenko, "Theoretical and empirical analysis of ReliefF and RReliefF," *Mach. Learn.*, vol. 53, no. 1/2, pp. 23–69, 2003.
- [78] A. G. Karegowda, A. S. Manjunath, and M. A. Jayaram, "Comparative study of attribute selection using gain ratio and correlation based feature selection," *Int. J. Inf. Technol. Knowl. Manage.*, vol. 2, no. 2, pp. 271–277, 2010.
- [79] H. Yang and J. Moody, "Feature selection based on joint mutual information," in *Proc. Int. ICSC Symp. Adv. Intell. Data Anal.*, 1999, pp. 22–25.
- [80] P. E. Meyer and G. Bontempi, "On the use of variable complementarity for feature selection in cancer classification," in *Applications of Evolutionary Computation*. Berlin, Springer: Springer, 2006, pp. 91–102.
- [81] A. Jakulin, "Machine learning based on attribute interactions," Doctoral dissertation, Univerza v Ljubljani, Ljubljana, Slovenia, 2005.
- [82] W. Shang, H. Huang, H. Zhu, Y. Lin, Y. Qu, and Z. Wang, "A novel feature selection algorithm for text categorization," *Expert Syst. Appl.*, vol. 33, no. 1, pp. 1–5, 2007.
- [83] Q. Gu, Z. Li, and J. Han, "Generalized fisher score for feature selection," 2012, *arXiv:1202.3725*.
- [84] K. Torkkola, "Feature extraction by non-parametric mutual information maximization," *J. Mach. Learn. Res.*, vol. 3, pp. 1415–1438, 2003.
- [85] F. Fleuret, "Fast binary feature selection with conditional mutual information," *J. Mach. Learn. Res.*, vol. 5, pp. 1531–1555, 2004.
- [86] D. Lin and X. Tang, "Conditional infomax learning: An integrated framework for feature extraction and fusion," in *Proc. Eur. Conf. Comput. Vis.*, 2006, pp. 68–82.
- [87] F. Nie, S. Xiang, Y. Jia, C. Zhang, and S. Yan, "Trace ratio criterion for feature selection," *AAAI*, vol. 2, pp. 671–676, 2008.
- [88] L. Prokhorenkova, G. Gusev, A. Vorobev, A. V. Dorogush, and A. Gulin, "CatBoost: Unbiased boosting with categorical features," in *Proc. Neural Inf. Process. Syst. Conf.*, 2018, pp. 6638–6648.



**Alim Samat** (Member, IEEE) received the B.S. degree in geographic information system from Nanjing University, Nanjing, China, in 2009, the M.E. degree in photogrammetry and remote sensing from the China University of Mining and Technology, Xuzhou, China, in 2012, and the Ph.D. degree in cartography and geography information system from Nanjing University, in 2015.

He was a Visiting Ph.D. Student with the Department of Industrial and Information Engineering, University of Pavia, Pavia, Italy, in 2014. He is currently an Associate Professor with the State Key Laboratory of Desert and Oasis Ecology, Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, Ürümqi, China. His main research interests include optical and PolSAR image processing and applications, machine learning, and pattern recognition.

Dr. Samat is a Reviewer for several international remote sensing journals, including *Remote Sensing Environment*, *IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING*, *IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATION AND REMOTE SENSING*, *IEEE Geoscience and Remote Sensing Letters*, *Remote Sensing*, *International Journal of Remote Sensing*, etc.



**Erzhu Li** received the Ph.D. degree in cartography and geographic information system from Nanjing University, Nanjing, China, 2017.

He is currently a Lecturer and Researcher with the School of Geography, Geomatics and Planning, Jiangsu Normal University, Xuzhou, China. His research interests include high-resolution image processing and computer vision in urban remote sensing applications.

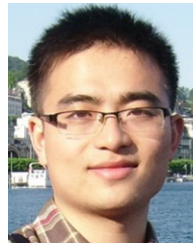
Dr. Li is a Reviewer for several international remote sensing journals, including the *IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING*, *IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATION AND REMOTE SENSING*, *IEEE Geoscience and Remote Sensing Letters*, *Remote Sensing*, etc.



**Peijun Du** (Senior Member, IEEE) received the Ph.D. degree in geodesy and survey engineering from the China University of Mining and Technology, Xuzhou, China, in 2001.

He is a Professor of remote sensing and geographical information science with Nanjing University, Nanjing, China. He has authored more than 70 articles in international peer-reviewed journals and more than 100 articles in international conferences and Chinese journals. His research interests include remote sensing image processing and pattern recognition, hyperspectral remote sensing, and applications of geospatial information technologies.

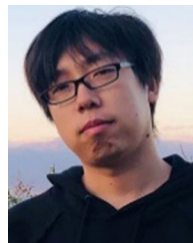
Dr. Du is an Associate Editor for the *IEEE Geoscience and Remote Sensing Letters*. He was the Co-Chair of the Technical Committee of URBAN in 2009, the International Association of Pattern Recognition Workshop on Pattern Recognition in Remote Sensing (IAPR-PRRS) in 2012, the International Workshop on Earth Observation and Remote Sensing Applications (EORSA) in 2014, the Co-Chair of the Local Organizing Committee of Joint Urban Remote Sensing Event (JURSE) in 2009, the Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing (WHISPERS) in 2012, and EORSA in 2012, and a member of the Scientific Committee or the Technical Committee of other international conferences, including WHISPERS from 2010 to 2016, URBAN in 2011, 2013, and 2015, MultiTemp in 2011, 2013, and 2015, the International Symposium on Image and Data Fusion (ISIDF) in 2011, and the International Society for Optics and Photonics (SPIE) European Conference on Image and Signal Processing for Remote Sensing from 2012 to 2016.



**Sicong Liu** (Member, IEEE) received the B.Sc. degree in geographical information system and the M.E. degree in photogrammetry and remote sensing from the China University of Mining and Technology, Xuzhou, China, in 2009 and 2011, respectively, and the Ph.D. degree in information and communication technology from the University of Trento, Trento, Italy, 2015.

He is currently an Associate Professor with the College of Surveying and Geo-Informatics, Tongji University, Shanghai, China. His research interests include multitemporal data processing, change detection, and spectral signal analysis in multispectral/hyperspectral images.

Dr. Liu is a Reviewer for several international remote sensing journals, including the *IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING*, *IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATION AND REMOTE SENSING*, *IEEE Geoscience and Remote Sensing Letters*, *ISPRS Journal of Photogrammetry and Remote Sensing*, *Remote Sensing*, etc.



**Junshi Xia** (Senior Member, IEEE) received the B.S. degree in geographic information systems and the Ph.D. degree in photogrammetry and remote sensing from the China University of Mining and Technology, Xuzhou, China, in 2008 and 2013, respectively, and the Ph.D. degree in image processing from the Grenoble Images Speech Signals and Automatics Laboratory, Grenoble Institute of Technology, Grenoble, France, in 2014.

From 2014 to 2015, he was a Visiting Scientist with the Department of Geographic Information Sciences, Nanjing University, Nanjing, China. From 2015 to 2016, he was a Postdoctoral Research Fellow with the University of Bordeaux, Bordeaux, France. From 2016 to 2018, he was the Japan Society for the Promotion of Science Postdoctoral Overseas Research Fellow with the University of Tokyo, Tokyo, Japan. Since 2018, he has been a Research Scientist with RIKEN Center for Advanced Intelligence Project (AIP), Tokyo, Japan. His research interests include multiple classifier systems in remote sensing, hyperspectral remote sensing image processing, and deep learning in remote sensing applications.

Dr. Xia was the recipient of the first place prize in the *IEEE Geoscience and Remote Sensing Society Data Fusion Contest* organized by the Image Analysis and Data Fusion Technical Committee in 2017. Since 2019, he has been an Associate Editor for the *IEEE GEOSCIENCE AND REMOTE SENSING LETTERS (GRSL)*, and Guest Editor for *Remote Sensing* and the *IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING (JSTARS)*.