Predicting Classification Performance for Benchmark Hyperspectral Datasets

Bin Zhao, *Member, IEEE*, Haukur Isfeld Ragnarsson, *Student Member, IEEE*, Magnus O. Ulfarsson, *Senior Member, IEEE*, Gabriele Cavallaro, *Member, IEEE*, and Jón Atli Benediktsson, *Fellow, IEEE*

***Abstract*—The classification of hyperspectral images (HSIs) is an essential application of remote sensing and it is addressed by numerous publications every year. A large body of these papers present new classification algorithms and benchmark them against established methods on public hyperspectral datasets. The metadata contained in these research papers (i.e., the size of the image, the number of classes, the type of classifier, etc.) present an unexploited source of information that can be used to estimate the performance of classifiers before doing the actual experiments. In this article, we propose a novel approach to investigate to what degree HSIs can be classified by using only metadata. This can guide remote sensing researchers to identify optimal classifiers and develop new algorithms. In the experiments, different linear and nonlinear prediction methods are trained and tested by using data on classification accuracy and metadata from 100 HSIs classification papers. The experimental results demonstrate that the proposed ensemble learning voting method outperforms other comparative methods in quantitative assessments.**

***Index Terms*—Hyperspectral image (HSI) classification, prediction, remote sensing.**

I. Introduction

HYPERSPECTRAL remote sensing images have hundreds of spectral bands that cover a continuous range of wavelengths from shortwave infrared to visible range [1]. Imaging scenes usually have many different materials that have different reflectance at different wavelengths. Therefore, each material has a unique spectral signature that can be detected from the hyperspectral images (HSIs). Hyperspectral imaging has application in various fields, such as agriculture [2], mineralogy [3], and surveillance [4].

HSIs classification is the process of transforming HSIs into land-cover maps and is a fundamental application of hyperspectral imaging. Numerous HSI classification methods have been developed in the last decades, ranging from spectral methods [5], spectral–spatial methods [6], to deep learning methods [7], etc. HSI classification is a complex process and the accuracy of its results depends on the properties of the input data, the type of classification algorithm, and other metadata containing features that describe the classification experiment. Considerable effort has gone into comparing such HSI classification methods. The most common approach is to benchmark the algorithm against established classifiers on public datasets, while keeping other experimental features, such as the sampling procedure of the training and test set, fixed between experiments [8], [9]. In [10], statistical meta-analysis is employed to compare pairs of classification algorithms. In [11], scatterplots are used to investigate the relationship between classification accuracy and various metadata features.

Research papers that propose novel HSI classification methods include several pieces of information (i.e., metadata) that describe the experiments. This raises the question to what degree the metadata can predict the accuracy of a classification algorithm. To help answer this question, we develop a novel method to predict the classification accuracy based on features that are typically reported in classification experiments. This approach can support the remote sensing community by guiding its researchers and reviewers to undertake new ideas. The proposed prediction method can be advantageous in applied classification problems as a baseline method.

To develop the predictor, we analyze one hundred papers on HSI classification with a combination of processing techniques and classification algorithms. Fig. 1 reports the number of papers that are sampled in each year and the type of classifiers. Each article presents at least one new classification method and reports performance metrics for the proposed methods and comparison methods. The total number of reported performance metrics is 2964 from 686 classification methods. Some papers [9], [12]–[14] are overview papers that cover a wide range of known methods, while others [15], [16] present novel classifiers to the problem. The applied techniques and algorithms have a significant impact on the outcome. They can be divided into categories as presented in [9]. These categories are used to sample data from ieeexplore.org.[1] To restrict the search for papers to remote sensing, the papers are searched on hyperspectral classification and used keywords for each category to find relevant papers. For example, papers on mathematical morphology-based classification (MMC) were found with the search query "remote sensing," "hyperspectral classification," and "morphology." For each set of results, we sample few of the most cited papers and some additional papers that are more recent or more relevant.

Manuscript received March 23, 2022; revised April 24, 2022; accepted May 6, 2022. Date of publication May 10, 2022; date of current version June 2, 2022. This work was supported in part by the Research Fund of the University of Iceland. *(Corresponding author: Jón Atli Benediktsson.)*

Bin Zhao, Haukur Isfeld Ragnarsson, Magnus O. Ulfarsson, and Jón Atli Benediktsson are with the Faculty of Electrical and Computer Engineering, University of Iceland, 101 Reykjavik, Iceland (e-mail: biz1@hi.is; haukurisfeld@gmail.com; mou@hi.is; benedikt@hi.is).

Gabriele Cavallaro is with Jülich Supercomputing Centre, 52428 Jülich, Germany (e-mail: g.cavallaro@fz-juelich.de).

Digital Object Identifier 10.1109/JSTARS.2022.3173893

[1][Online]. Available: https://ieeexplore.ieee.org/Xplore/home.jsp

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

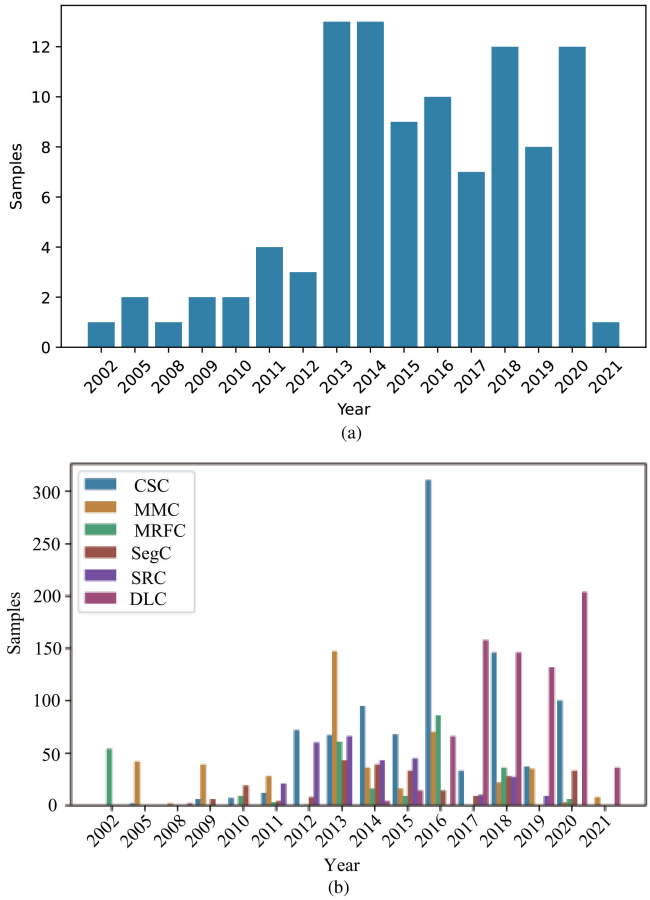


Fig. 1. (a) Overall sampling of papers by year. (b) Overall sampling of methods by year, split by method category.

Additionally, papers that have open source code available were found by searching code repositories such as github.com[2] for the term "hyperspectral-image-classification." This search yielded several repositories and relevant papers.

The rest of this article is organized as follows. Section II illustrates the proposed prediction methods. Section III introduces the 17 types of different features in the metedata extracting from the 100 HSI classification papers used for the prediction method. Section IV describes the different types of HSI classification methods used in the 100 HSI classification papers that we utilize to develop our prediction method. The experimental results are given in Section V. Finally, Section VI concludes the article.

II. Prediction Methods

Two prediction methods are investigated in this article. The first method is the ordinary least squares method (OrdLS), which is a linear method. The second method is the ensemble learning voting method (EnsLV), which is a nonlinear method. The metadata, including 17 types of different features extracting from the 100 HSI classification papers, is used for the training and test sets for the two prediction methods. A detailed description of the features in the metadata is given in Section III.

[2][Online]. Available: https://github.com/

TABLE I
Summary of Features

Feature	Label	Encoding
Number of training samples	Ntr	Integer > 0
Number of test samples	Nts	Integer > 0
Number of classes	Ncl	Integer > 0
Number of spectral bands	Nsb	Integer > 0
Resolution	Res	Float > 0.0
Scene	Scn	Binary
Sampling method	Split	Binary
Dimensionality reduction	DR	Binary
Method type	MT	Binary
Method category	MC	Treatment
Classical spectral-based classifier (CSC)	MC1	Binary
Mathematical morphology-based classifier (MMC)	MC2	Binary
Markov random fields-based classifier (MRFC)	MC3	Binary
Segmentation-based classifier (SegC)	MC4	Binary
Sparse representation-based classifier (SRC)	MC5	Binary
Deep learning-based classifier (DLC)	MC6	Binary
Averaging	Avg	Integer > 0
Open source code	OSCode	Binary

A. OrdLS

OrdLS used for regression analysis is an ordinary least squares regression. The regression model is given by

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \tag{1}$$

where $\boldsymbol{\epsilon}$ is white zero-mean Gaussian noise with variance σ^2

$$\mathbf{y} = [y_1, \ldots, y_n]^T$$
$$\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_n]^T$$
$$\mathbf{x}_i = [1, \mathrm{Ntr}_i, \mathrm{Nts}_i, \mathrm{Ncl}_i, \mathrm{Nsb}_i, \mathrm{Res}_i, \mathrm{Scn}_i, \mathrm{Split}_i, \mathrm{DR}_i, \mathrm{MT}_i, \mathrm{MC2}_i, \mathrm{MC3}_i, \mathrm{MC4}_i, \mathrm{MC5}_i, \mathrm{MC6}_i, \mathrm{Avg}_i, \mathrm{OSCode}_i]^T$$
$$\boldsymbol{\beta} = [\beta_0, \ldots, \beta_{17}]^T$$

where $\mathbf{y}$ is the target variable, $\mathbf{x}_i$ contains the metadata along with an intercept as the first element for sample i, Table I defines the metadata features, n is the number of training samples, and $\boldsymbol{\beta}$ are the regression parameters. For model (1), the maximum likelihood estimate is given by

$$\widehat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}. \tag{2}$$

The target variables $\boldsymbol{y}$ used in the article are overall accuracy (OA), average accuracy (AA), and Kappa coefficient (Kappa). OA is the percentage of accurately classified samples over the entire test set. AA is the average of each class's percentage of accurately classified samples. Kappa is a measure of how OA compares to expected accuracy.

To further analyze the effectiveness of OrdLS, some penalties can be added to OrdLS; thus, an extension study of OrdLS is performed. OrdLS is compared with the following prediction methods.

1) *Orthogonal Matching Pursuit (OMP)*: The ℓ_0 penalty is added to the regression model, i.e., the cost function is

$$\widehat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|_F^2, \text{ s.t. } \|\boldsymbol{\beta}\|_0 < k. \tag{3}$$

TABLE II
BREAKDOWN OF HYPERSPECTRAL IMAGES

Image	Sensor	No. of classes	No. of spectral bands	Resolution (m)	No. of experiments	OA	AA	Kappa
Indian Pines	AVIRIS	16	220	20	816	83.54±13.2	83.77±14.2	0.82±0.15
Salinas	AVIRIS	16	220	3.7	317	92.12±6.73	92.12±9.09	0.91±0.08
Salinas A	AVIRIS	6	220	3.7	23	94.00±6.20	91.86±7.34	0.89±0.11
Pavia University	ROSIS	9	103	1.3	973	86.46±10.4	88.97±9.62	0.83±0.14
Pavia Center	ROSIS	9	103	1.3	290	95.14±5.38	93.79±6.48	0.96±0.02
Botswana	Hyperion	14	145	30	94	92.53±5.58	95.62±4.78	0.95±0.05
Houston	NCALM	15	144	2.5	81	79.36±10.3	80.69±10.4	0.77±0.11
Kennedy Space Center	AVIRIS	13	176	18	209	89.41±9.69	85.14±12.31	0.87±0.10
Washington DC Mall	HYDICE	6	191	2.5	161	92.57±6.59	91.44±5.24	0.93±0.04

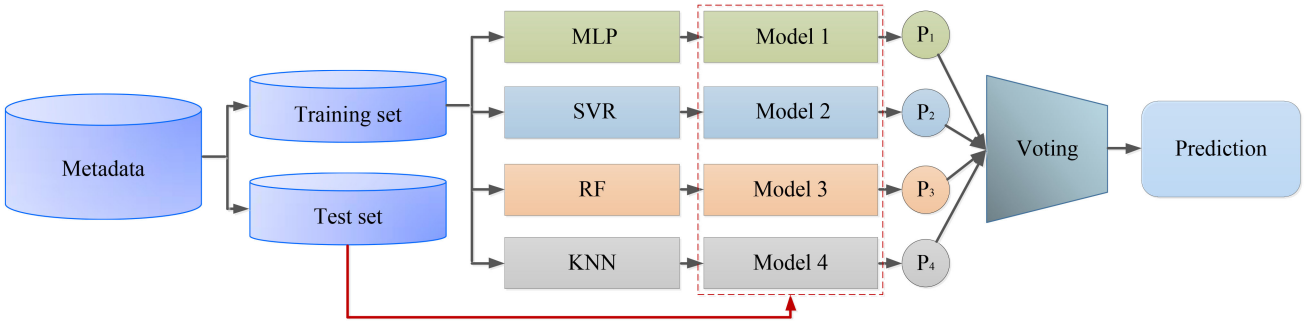


Fig. 2. Schematic of the EnsLV.

TABLE III
BREAKDOWN OF CLASSIFICATION METHOD CATEGORIES

Category	Used times	OA	AA	Kappa
CSC	956	85.25±11.34	85.17±12.37	0.85±0.14
MMC	448	87.98±11.10	89.13±8.14	0.85±0.12
MRFC	281	88.04±8.490	90.01±4.33	0.90±0.05
SegC	236	86.53±10.62	87.26±13.24	0.85±0.13
SRC	281	88.25±11.00	88.74±8.910	0.87±0.12
DLC	762	90.64±10.86	89.81±12.96	0.87±0.14

2) *Ridge regression (RigR)*: The ℓ_2 penalty is added to the regression model, i.e., the cost function is

$$\widehat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \ \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|_F^2 + \lambda_1 \|\boldsymbol{\beta}\|_F^2. \tag{4}$$

3) *Lasso regression (LasR)*: The ℓ_1 penalty is added to the regression model, i.e., the cost function is

$$\widehat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \ \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|_F^2 + \lambda_2 \|\boldsymbol{\beta}\|_1. \tag{5}$$

4) *Elastic net (EltN)*: The ℓ_2 and ℓ_1 penalties are added to the regression model, i.e., the cost function is

$$\widehat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \ \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|_F^2 + \lambda_3 \|\boldsymbol{\beta}\|_1 + \lambda_4 \|\boldsymbol{\beta}\|_F^2. \tag{6}$$

Besides the aforementioned methods, the mean of the predicted variables in the training dataset is defined as a baseline method to be compared. The baseline method is used to evaluate the performance of the prediction methods. Any other prediction methods should at least perform equally well and, ideally, better than the baseline method.

Table V presents the prediction results (detailed experimental setting is given in Section V). OrdLS performs similarly as the penalized least squares methods 1)–4), and all of them perform better than the baseline prediction method. Thus, the penalty is not needed for the linear prediction.

TABLE IV
RESULTS OF REGRESSION ANALYSIS

	OA		AA		Kappa	
	MAE	R^2	MAE	R^2	MAE	R^2
	0.074	0.183	0.073	0.227	0.091	0.194
Feature	T	P-value	T	P-value	T	P-value
Intercept	119.67	0.000	99.783	0.000	77.582	0.000
Ntr	10.946	0.000	8.5100	0.000	10.989	0.000
Nts	−5.6740	0.000	−4.0840	0.000	−2.0360	0.042
Ncl	−7.1140	0.000	−5.7860	0.000	−4.2990	0.000
Nsb	2.6390	0.008	−1.7960	0.073	5.2800	0.000
Res	−0.7180	0.473	−2.2580	0.024	0.2660	0.791
Scn	−1.0270	0.305	−5.2590	0.000	0.0960	0.923
Split	−0.4000	0.689	5.7980	0.000	4.8750	0.000
DR	1.5750	0.115	−0.4350	0.663	−2.0320	0.042
MT	13.074	0.000	12.004	0.000	9.1230	0.000
MC2	3.1450	0.002	1.0310	0.303	0.5550	0.579
MC3	3.3650	0.001	−0.0210	0.983	1.3260	0.185
MC4	1.8170	0.069	−0.1620	0.871	0.1670	0.868
MC5	6.7590	0.000	5.2440	0.000	5.8020	0.000
MC6	5.2280	0.000	2.2970	0.022	2.4180	0.016
Avg	1.0190	0.308	2.2870	0.022	1.6000	0.110
OSCode	2.2410	0.025	3.2480	0.001	2.7630	0.006

B. EnsLV

EnsLV uses and combines several prediction models for prediction to improve the robustness over a single estimator. The mutilayer perceptron (MLP), support vector regression (SVR), random forest (RF), K-nearest neighbors (KNN), adaboost (AdaT), and histogram gradient boosting (HGB) nonlinear prediction methods are selected as the based-estimators to train EnsLV. Auto-sklearn [17] and Bayesian optimization

TABLE V
OVERALL ACCURACY PREDICTION RESULTS FOR THE LINEAR REGRESSION MODELS

Method	Transformations					
	No	Logit	Standard	Power	Gaussian	Uniform
Baseline	8.46 ± 0.34	8.06 ± 0.34	8.46 ± 0.34	8.08 ± 0.35	8.03 ± 0.35	8.03 ± 0.35
RigR	$\mathbf{7.43 \pm 0.33}$	6.97 ± 0.31	$\mathbf{7.43 \pm 0.33}$	7.03 ± 0.34	6.93 ± 0.32	7.02 ± 0.34
LasR	7.74 ± 0.38	7.20 ± 0.36	7.74 ± 0.38	7.39 ± 0.36	7.20 ± 0.35	7.36 ± 0.34
OMP	7.78 ± 0.41	7.23 ± 0.38	7.78 ± 0.41	7.39 ± 0.37	7.22 ± 0.38	7.32 ± 0.36
EltN	7.74 ± 0.38	7.21 ± 0.34	7.75 ± 0.38	7.38 ± 0.35	7.21 ± 0.35	7.36 ± 0.34
OrdLS	$\mathbf{7.43 \pm 0.33}$	$\mathbf{6.96 \pm 0.30}$	$\mathbf{7.43 \pm 0.33}$	$\mathbf{7.02 \pm 0.34}$	$\mathbf{6.92 \pm 0.32}$	$\mathbf{7.01 \pm 0.33}$
Average	7.62 ± 0.37	7.11 ± 0.34	7.63 ± 0.37	7.24 ± 0.35	7.10 ± 0.34	7.21 ± 0.34

Values are given as mean absolute error (%) with standard deviation. The best results are in bold typeface

TABLE VI
OVERALL ACCURACY PREDICTION RESULTS FOR THE NONLINEAR REGRESSION MODELS

Method	Transformations					
	No	Logit	Standard	Power	Gaussian	Uniform
SVR	5.32 ± 0.19	5.17 ± 0.22	5.22 ± 0.20	5.44 ± 0.19	5.14 ± 0.21	5.19 ± 0.25
MLP	4.62 ± 0.16	4.39 ± 0.12	4.38 ± 0.09	4.29 ± 0.12	4.30 ± 0.21	4.67 ± 0.26
RF	4.00 ± 0.18	3.88 ± 0.15	4.00 ± 0.18	3.89 ± 0.16	3.91 ± 0.17	3.89 ± 0.16
AdaT	4.28 ± 0.17	4.17 ± 0.14	4.28 ± 0.18	4.16 ± 0.12	4.23 ± 0.19	4.16 ± 0.13
HGB	4.21 ± 0.18	4.19 ± 0.15	4.21 ± 0.18	4.25 ± 0.14	4.10 ± 0.17	4.39 ± 0.16
KNN	4.12 ± 0.21	4.02 ± 0.21	4.12 ± 0.21	4.02 ± 0.19	4.04 ± 0.22	4.01 ± 0.19
EnsLV	$\mathbf{3.93 \pm 0.17}$	$\mathbf{3.82 \pm 0.18}$	$\mathbf{3.93 \pm 0.18}$	$\mathbf{3.84 \pm 0.18}$	$\mathbf{3.85 \pm 0.19}$	$\mathbf{3.84 \pm 0.18}$
Average	4.43 ± 0.18	4.30 ± 0.17	4.37 ± 0.17	4.34 ± 0.15	4.29 ± 0.20	4.39 ± 0.19

Values are given as mean absolute error (%) with standard deviation. The best results are in bold typeface

are used to find the best ensemble of the nonlinear regression methods for EnsLV and the soft voting is used for EnsLV to perform prediction. The EnsLV method is illustrated in Fig. 2.

III. FEATURES IN METADATA

The 17 types of different features in the metadata extracting from the 100 HSI classification papers used for the prediction methods are summarized in Table I and are described as follows.

A. Dataset Features

Much information of the experimental datasets used in the 100 HSIs classification papers can be extracted and used as features for the prediction methods. The number of classes in the dataset (Ncl), the number of spectral bands in the classification dataset (Nsb), the resolution of the dataset (Res), and the type of scene (whether it is rural or urban) depicted in the datasets (Scn) are used as features for the prediction models.

B. Sampling Methods

This article focuses on supervised classification. The classifier model is trained using the labeled training samples and then is evaluated using the labeled test samples. The training samples and the test samples are selected from the entire labeled samples pool of the experimental dataset. There are two main ways of selecting the training samples. The first is to randomly select the training samples from the entire labeled samples pool. The second is to select the training samples randomly or by hand, a certain percentage or a fixed number, from each labeled class from the entire labeled samples pool. The latter way ensures that the training set includes samples from all the classes of the experimental dataset. The remaining samples of the entire labeled samples pool are selected as test samples. Thus, the type of selecting training samples (Split), the number of training samples (Ntr), and the test samples (Nts) are used as features for the prediction models.

C. Dimensionality Reduction

HSIs usually have hundreds of spectral bands and provide abundant spectral information about a scene. However, the high dimensionality of HSIs makes the processing computationally and memory costly. To achieve an acceptable classification accuracy for a high-dimensional image many conventional HSI processing methods require many training samples. This is known as the Hughes phenomenon or the curse of dimensionality. Thus, when the number of training samples is limited, dimensionality reduction (DR) is selected to be a tradeoff way between classification accuracy and the number of spectral bands to solve this problem. Dimensionality reduced data should be a good representation of the original data. In addition, both the computing time and the number of training samples required will become less when the data dimensionality is lower. Therefore, DR is a very important preprocessing step for HSI classification. Whether the DR has been done for the experimental dataset or not is used as the feature for the prediction models. Sometimes, the DR is embedded in the classification algorithms, and it is difficult to determine what number of the spectral bands are used for the real classification. In these cases, the number of spectral bands is set equal to the original dimension of the experimental dataset.

D. Algorithm Categories

Most HSI classification methods fall into either the spectral or the spectral–spatial method category. Spectral methods utilize only spectral information for the pixel being classified, while spectral–spatial methods also incorporate spatial information

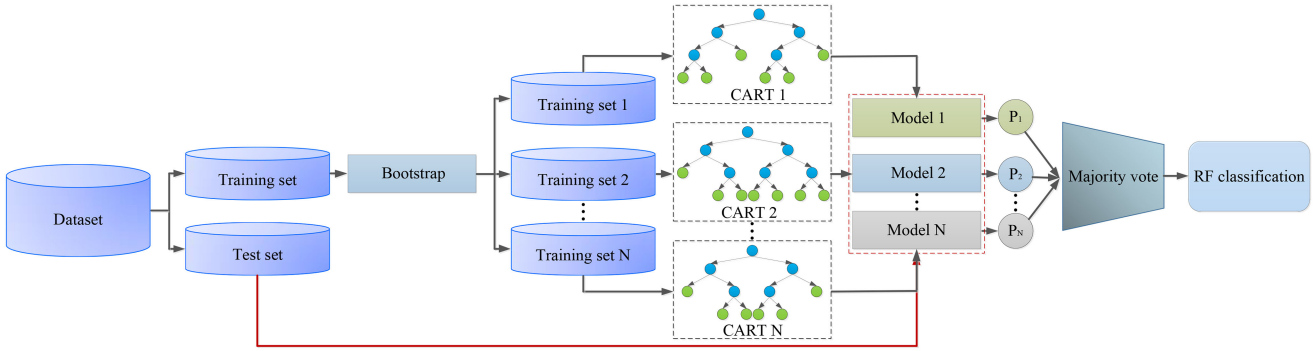


Fig. 3. Schematic of the RF.

from the surrounding pixels. Therefore, the type (spectral or spectral–spatial) of the HSI classification methods (MT) is used as a feature for the prediction models. In addition, we consider here six more specific categories for HSI classification methods (MC) as additional features for the prediction models (see Section IV for further information).

E. Algorithm Reliability

In addition to the aforementioned features, two other features are considered for the prediction models. The first is the number of experiments performed for averaging (Avg) of the result, whether by cross-validation or not. The second feature is whether the code of the classification algorithm is open-sourced or somehow available for evaluation (OSCode). Both features indicate whether the result is adequately validated and how reliable it is.

IV. Classification Methods

In this section, we describe the methods that are used in the 100 HSI classification papers that we use to develop our prediction method. The methods are categorized into six groups: classical spectral-based classifier (CSC), mathematical morphology-based classifier (MMC), Markov random fields-based classifier (MRFC), segmentation-based classifier (SegC), sparse representation-based classifier (SRC), and deep learning based classifier (DLC). The category of classification methods (MC) is used as a feature for the prediction models. The six groups are denoted as MC1 to MC6.

A. CSC

CSCs, such as support vector machines (SVMs) [18] and RF [19], are usually considered as baseline classification methods in classification paper. For an image pixel (spectral vector) that belongs to a certain class, the SVM classifier tries to find a linearly separating hyperplane between different classes such that the margin between the hyperplane and the closest training samples is maximized. In real HSI applications, linear separability is usually not satisfied. Therefore, a kernel method is utilized to map the data using a nonlinear transformation to a higher dimensional space and, in that space, try to find a linearly separating hyperplane between different classes.

RF is another method commonly used in HSI classification that uses spectral vector as an unit of analysis. RF is an ensemble learning method that uses decision tree as a base classifier and combines their decisions by using a majority vote for final classification. The bootstrap aggregating (bagging) method is used for RF to randomly select subset sampled from the entire training dataset to construct the decision trees, which also acts to prevent the problem of overfitting by individual decision trees and reduces the classification error. The classification and regression tree is used to train each decision tree. The predicted class of an observation is calculated based on the majority vote of the decision trees. The schematic of the RF is shown in Fig. 3.

B. MMC

MMC methods use mathematical morphology (MM) [14], [20] to analyze the spatial information of image for HSI classification. Two fundamental MM operations are dilation and erosion [14], [20]. Dilation adds pixels to the boundaries of objects in an image, while erosion removes pixels on object boundaries. The number and position of pixels added or removed from the objects in an image depends on the size and shape of the structuring element (SE) used to process the image. Thus, the SE is used for MM to extract or suppress HSI structures by checking that each position of the SE fits within the image objects. All other MM operations can be expressed by erosion and dilation [21]. For example, the morphological opening operation is to dilate an eroded image to filter out bright structures, while the morphological closing operation is to erode a dilated image to suppress dark structures. The reconstruction filter is implemented on these operators to preserve original structures of image and suppress shape noise. The morphological opening and closing operations can be effectively used for analyzing and processing spatial information for image classification.

Morphological profiles (MPs) [22] are a classical example of MM method, which are constructed based on opening and closing operators from a grayscale image using SEs with increasing sizes, produced the multiscale spatial information of the image, and improved the traditional pixelwise spectral classification. Attribute profiles (APs) [23] are another classical example of MM method, which can extract spatial and contextual features based on multiple attributes from a grayscale image. MPs and APs are used for grayscale images. Based on MPs and APs,

extended MP (EMP) [22] and extended AP (EAP) [24] are developed for dealing with HSI. The spectral dimensionality of HSI is first reduced by using principal component analysis (PCA). EMP and EAP construct MPs and APs, respectively, on the first few principal components (PCs). Both EMP and EAP used for HSI classification exhibit good performance [25]. It should be noted that the key issue for the performance of MMC methods is to determine a set of base images, on which the EMP and EAP are built. The base images are feature images extracted from the original HSI based on feature extraction for HSI classification. Thus, many feature extraction methods can be used to improve MMC classification performance, such as Fauvel *et al.* [26] used kernel PCA to extract the base images for EMP to improve the HSI classification performance. The paper [23] proved that independent component analysis is more suitable than PCA for generating the base images for EAP. More developed MMC methods are described in [9], [27]–[43].

C. MRFC

MRFC combines spatial neighborhood information into the posterior probability of the spectral features for image classification. Based on the Hammersly–Clifford theorem [44], the global posterior probability distribution of MRFC follows a Gibbs distribution and can be denoted as a function based on the spectral energy function and the spatial energy function. The spectral energy function is the cost of a pixel being assigned with different classes. The larger the probability of a pixel belonging to a certain class, the more probable that the pixel is assigned the corresponding label. The spatial energy function inspires the labels of neighboring pixels to be the same. Different MRFC methods can be constructed based on different spectral energy functions, such as the classic-MRFC method [44] used Euclidean distance for spectral energy function, the Mahalanobis distance was used for the Mahalanobis-MRFC method [45], and the spectral angle was used for the spectral angle-MRFC [46]. These MRFC methods can also combine with other classifiers such as SVM and Gaussian mixture models as shown in [47], belief propagations or tree reweighted message passing as shown in [48], convolutional neural networks (CNNs) as shown in [49], and more developed MRFC methods as described in [6], [50]–[60].

D. SegC

SegCs use segmentation algorithms to partition an image into nonoverlapping homogenous regions. Some algorithms, such as *K*-means clustering, watershed edge detection, and binary partition trees, can be used for SegC methods to segment an HSI to segmentation map. Each segmented region in the segmentation map can be viewed as a connected spatial neighborhood for all the pixels in the region. SegCs can use any pixelwise classifier and majority vote for each segmented region in the segmentation map, and all pixels in the same segmented region are labeled to the most frequent class with this region based on the pixelwise classification results [9]. Several SegC methods have been proposed for HSI classification. Ghamisi *et al.* [61] used the fractional-order Darwinian particle swarm optimization method for HSI segmentation and then used spectral classifier for HSI classification. Zhao *et al.* [62] used the optimal *K*-means clustering method for HSI segmentation and then used SVM classifier for image classification. Tarabalka *et al.* [63] used the ISODATA algorithm and Gaussian mixture resolving techniques for image clustering and then achieved classification map by performing a majority voting on the pixelwise SVM classification using adaptive neighborhoods defined by the segmentation map. More developed SegC methods are described in [64]–[69].

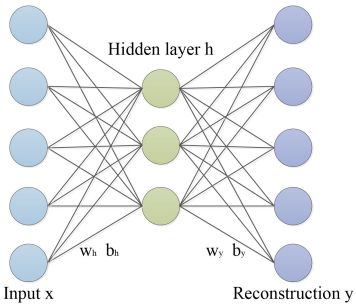


Fig. 4. Schematic of the AE.

E. SRC

Sparse representation is the method of representing a signal where only a few parameters are nonzero. SRC is a coding scheme that represents a signal by a sparse linear combination of samples (atoms) from a dictionary. Many SRC methods have been proposed for HSI classification. Chen *et al.* [70] proposed a SRC method based on the observation that the pixel can be represented by a sparse linear combination of atoms from a structured dictionary. Gao *et al.* [71] used a low-rank based sparse representation to generate high-quality HSIs yielding significant classification performance improvement. Fang *et al.* [72] used a multiscale adaptive sparse representation methods to improve the performance of HSI classification. More developed SRC methods are described in [16], [73]–[80].

F. DLC

DLC methods use deep neural networks to automatically learn deep features from data for HSI classification. Four representative methods, i.e., autoencoders (AEs), CNNs, recurrent neural networks (RNNs), and generative adversarial networks (GANs), for HSI classification can be categorized from DLC methods and are respectively described, and the relevant references are reviewed [15], [34], [49], [81]–[102].

1) AEs: AEs are designed to learn efficient representation of the input data. Fig. 4 shows the schematic of the AE. An AE consists of two parts. The first is an encoder, which encodes the input $\mathbf{x}$ to the hidden layer $\mathbf{h}$. The second is decoder, which maps $\mathbf{h}$ to output layer (reconstruction) $\mathbf{y}$. The AE network is trained by minimizing the reconstruction loss between input $\mathbf{x}$ and reconstruction $\mathbf{y}$. Thus, the learned features in hidden layer can be used for classification or used as the input of a higher layer to generate deeper features.

Numerous AE-based methods have been developed for HSI classification. Stacked AEs (SAEs) [7] is constructed by stacking the input and hidden layers of AEs together layer by layer and can be used for HSI classification. Chen *et al.* [7] developed SAE-LR method which combines PCA, SAEs, and logistic regression together to merge spectral and spatial features for

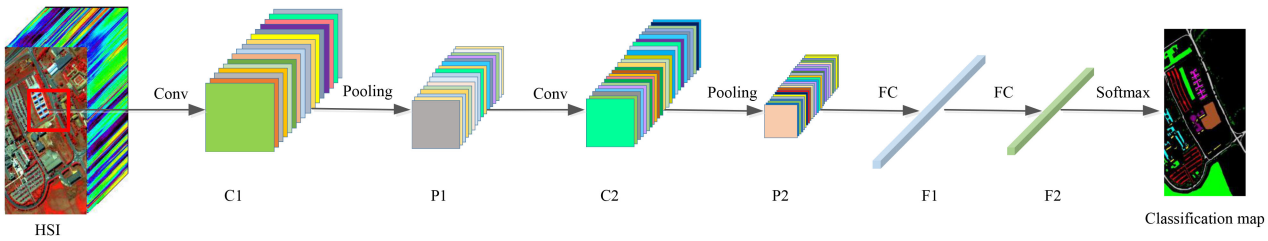


Fig. 5. Schematic of the CNN, which consists of two convolutional-pooling layers and two fully connected layers.

HSI classification. Xing *et al.* [103] utilized stacked denoise AE to pretrain the network and used the logistic regression method in the top layer of the network to perform supervised fine-tuning and classification. Some classification methods used AE to extract spectral features and utilized CNN to extract spatial features for HSI and has given good classification results [104], [105].

2) CNNs: A CNN is composed of stacked alternatively convolutional layers and pooling layers and ends with a fully connected neural networks and extracts the contextual 2-D spatial features of images by enforcing a local connectivity pattern between neurons of adjacent layers. Fig. 5 shows the framework of CNN. Convolutional layer is used for generating multiple feature maps based on convolution between the input dataset and the multiple learned filters. The rectified linear unit function is usually used for the activation function for CNN based on the advantages of low computation load, and robustness for gradient vanishing. Pooling operation is used for extracting more abstract features by reducing the spatial size of the feature maps. Max and average pooling are two common types of pooling operation. Max pooling takes the maximum value of each local cluster of neurons in the feature map, while average pooling extracts the average value. The fully connected layers are used to extract more deep and abstract features. The output from the final pooling layer is flattened and then acts as input to the fully connected layer. In order to achieve good classification performance for HSI, dropout is used for dealing with overfitting by setting the output of some hidden neurons to zero. Batch normalization is used to prevent gradient vanishing and speed up the training procedure.

Recently, numerous CNN-based methods have been developed for HSI classification. Some papers use DR for preprocessing in order to reduce the redundancy of spectral information of HSI and potentially mitigate the overfitting problem to improve CNN classification performance [106], [107]. Further, the HSI classification performance can be improved by combining with other techniques, such as in [108], Gabor filtering is used to extract spatial information, and in [109], sparse representation is used to refine the features learned from CNNs.

3) RNNs: RNN [110], [111] is an extension of the conventional feedforward neural network and can deal with sequences of data and dynamic temporal features by using a recurrent hidden state whose activation at each step depends on that of the previous steps. To deal with the vanishing gradient or exploding gradient for the long-term sequential data used for RNNs, long short-term memory [112], [113] and gated recurrent unit [93] can be used to deal with this problem. Considering each pixel vector in HSI as a set of orderly and continuing spectral sequences in the spectral space, RNNs can be effectively used for HSI classification. Paoletti *et al.* [114] used the simple recurrent unit to decouple the computational relationship between the current and previous states in the RNNs for HSI classification. In [115], pixel-matching and block-matching are used for calculating the similarity in RNNs for improving HSI classification performance.

4) GANs: GAN models contain a generative model G, which attempts to learn the distribution parameters from data and then generates new samples using the learned model for the purpose of generating fake inputs as real as possible, and a discriminative model D, which tries to the dependence of labels $\mathbf{y}$ on training data $\mathbf{x}$ and predicts $\mathbf{y}$ from $\mathbf{x}$ to distinguish between real and fake inputs. The training process of the discriminator will proceed both continuously and effectively by the adversarial manner and competition of the G and D models.

Recently, GAN-based methods have been effectively used for HSI classification. In [116], the GAN-based HSI classification framework, two CNNs are used to train generative model and discriminative model, respectively. A CNN was used for the generative model to generate the fake inputs, the other CNN was used to discriminate the inputs for the discriminative model. Dam *et al.* [117] developed multifake evolutionary based GAN method, where different generative objective losses are considered in the generator model to improve the classification performance of the discriminator model. In [116], 1-D GAN with spectral classifier and 3-D GAN with spectral–spatial classifier are developed for HSI classification.

V. Experimental Results

The experimental results of the proposed prediction methods based on the nine different real HSIs with considered 17 features are given in this section. Experimental results for OrdLS linear prediction method are compared with the following linear prediction methods: baseline, RigR, LasR, OMP, and EltN. Experimental results for EnsLV nonlinear prediction method are compared with the following nonlinear prediction methods: SVR, MLP, RF, AdaT, HGB, and KNN.

A. Datasets

We focus on datasets that have been used extensively in the literature. Therefore, nine real HSI datasets, Indian Pines, Salinas, Salinas A, Pavia University, Pavia Center, Bostwana, Houston, Kennedy Space Center, and Washington DC Mall datasets, used

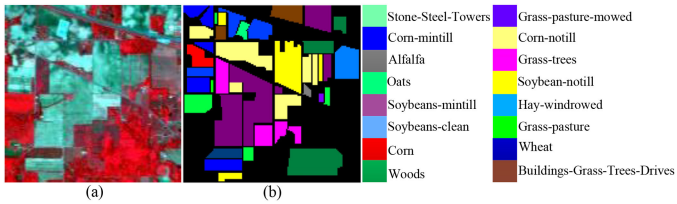


Fig. 6. Indian Pines dataset. (a) Three-band false-color image. (b) Ground-truth map.

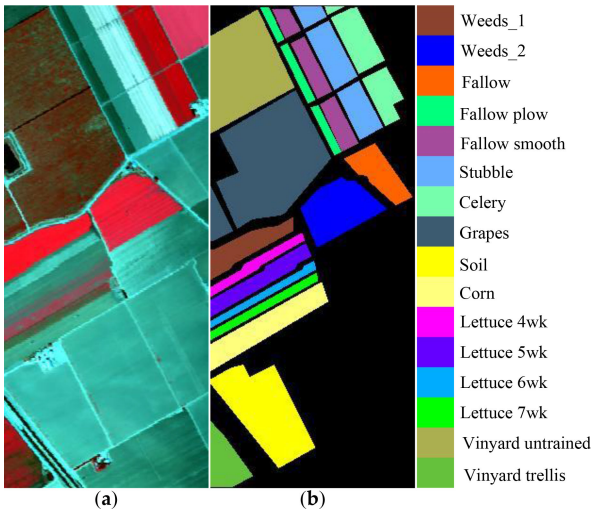


Fig. 7. Salinas dataset. (a) Three-band false-color image. (b) Ground-truth map reference.

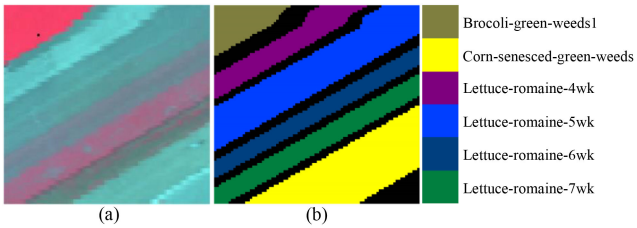


Fig. 8. Salinas A dataset. (a) Three-band false-color image. (b) Ground-truth map.

in the 100 HSIs classification papers are selected for the experimental datasets for the prediction methods. Table II provides detailed information for the nine HSI datasets, i.e., the sensor type, the number of classes in the dataset (Ncl), the number of spectral bands in the classification dataset (Nsb), and the spatial resolution of the dataset (Res). The number of experiments and the performance metrics, i.e., OA, AA, and Kappa, are collected and evaluated from the 100 HSI classification papers for the nine HSI datasets and are shown in Table II. Figs. 6– 14 show the false-color composite and the corresponding ground reference map for the nine HSI datasets, respectively. From Table II, it can be seen that the Houston dataset has the lowest OA, AA, and Kappa values, while, in contrast, the Pavia Centre has the highest values.

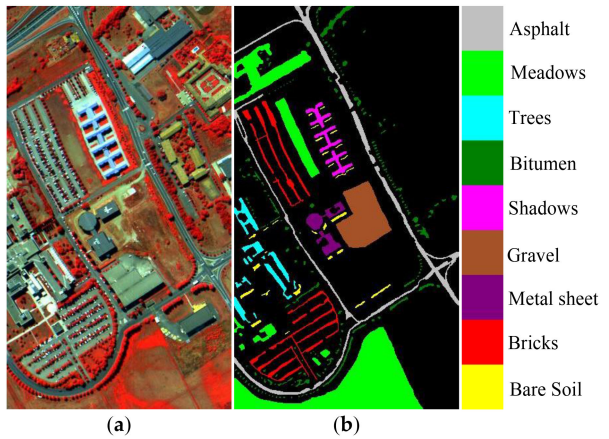


Fig. 9. Pavia University dataset. (a) Three-band false-color image. (b) Ground-truth map reference.

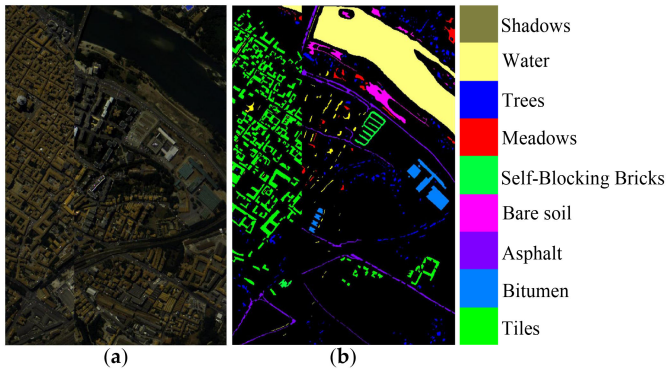


Fig. 10. Pavia Center dataset. (a) Three-band false-color image. (b) Ground-truth map reference.

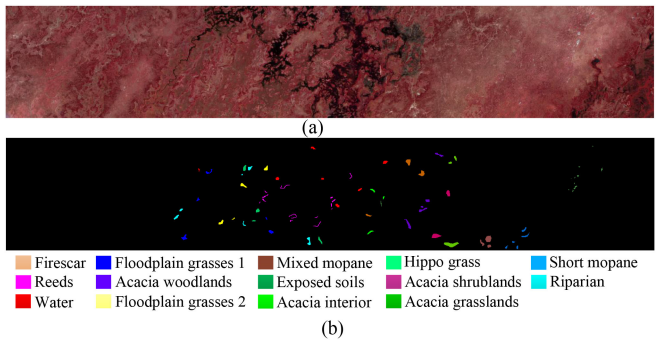


Fig. 11. Botswana dataset. (a) Three-band false-color image. (b) Ground truth-map reference.

Table III shows a breakdown of classification method categories. The CSCs, such as SVM and RF classifiers, are usually considered as baseline classification methods. It can be seen that the CSCs are the most heavily sampled and have the lowest OA, AA, and Kappa. Deep learning techniques have developed very fast recent years. The number of DLC methods is less than the CSCs and is much more than the other methods. DLCs have the highest OA. MRFCs have the highest AA and Kappa.

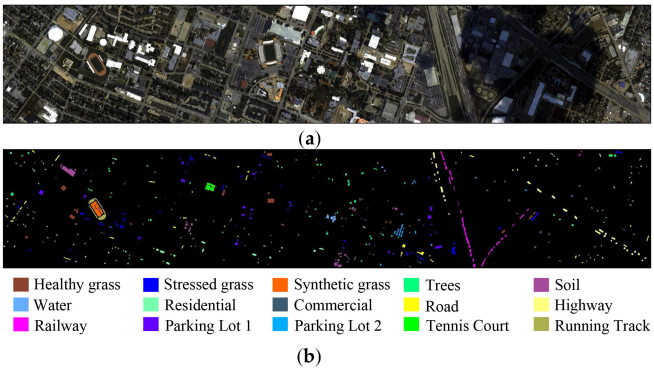


Fig. 12. Houston dataset. (a) Three-band false-color image. (b) Ground truth-map reference.

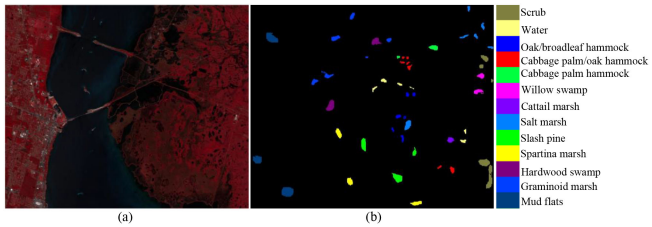


Fig. 13. Kennedy Space Center dataset. (a) Three-band false-color image. (b) Ground-truth map.

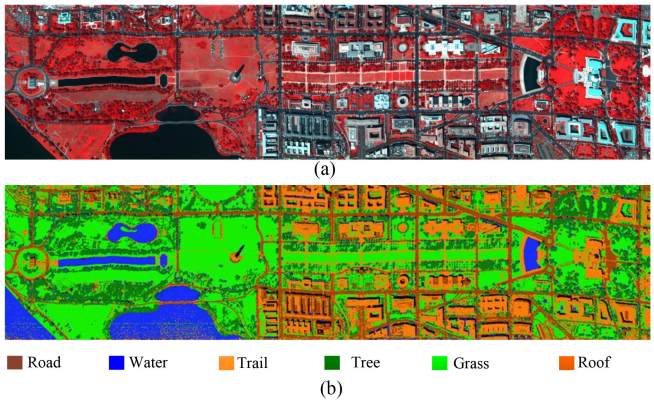
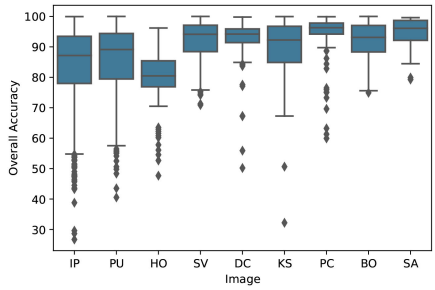
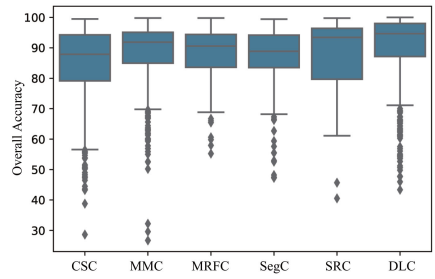


Fig. 14. Washington DC Mall dataset. (a) Three-band false-color image. (b) Ground-truth map reference.

Fig. 15(a) shows the boxplots for the distribution of the OA with respect to the nine different HSIs. It can be seen that how large the distribution is for the Indian Pines and Pavia University datasets. This is partly explained by the fact that they are the most heavily sampled in our datasets. It can also be seen how the mean accuracy is lowest for the Houston image. The Houston dataset is the second least sampled in our datasets and is also a very sparse urban scene which may be more difficult to classify. Fig. 15(b) shows the boxplots for the distribution of the OA with respect to the six different types of classifiers. It can be seen that deep learning classifiers have the highest mean accuracy.



(a) Image.



(b) Category.

Fig. 15. Distribution of overall accuracy with respect to (a) nine different HSIs and (b) six different types of classifiers.

B. Experimental Setup

The metadata with 17 different features are sample datasets extracted from the 100 HSI classification papers. The total number of samples in metadata is 2964. In the experiments, a certain amount of training samples are randomly selected from metadata for training the prediction models. The remaining samples in metadata are selected as test samples. All the parameters used for the comparison methods follow either the authors' source code settings or are experimentally evaluated according to the reference recommendations. It should be noted that EnsLV is an ensemble learning method that uses several based-estimators prediction models for predicting. The bootstrap aggregating method is used for EnsLV to randomly select a replaceable subset from the entire training dataset for training each based-estimator of EnsLV.

C. Evaluation Metrics

The quantitative metrics used to evaluate the results for the experimental datasets are the mean absolute error (MAE) between predicted and actual outcomes, the goodness of fit (R^2), the t-statistic, and the p-value. The definition of MAE is

$$\text{MAE} = \frac{\sum_{i=1}^{n} |y_i - \hat{y}_i|}{n} \tag{7}$$

where y_i is the observed value, $\hat{y}_i$ is the predicted value, and n is the number of samples. R^2 measures the coefficient of determination for linear prediction methods and is given by

$$R^2 = \frac{\sum(\hat{y}_i - \overline{y})^2}{\sum(y_i - \overline{y})^2} \tag{8}$$

where $\overline{y}$ is the mean of all observations. For perfect prediction, R^2 is equal to one. R^2 is a fraction of the variance of the

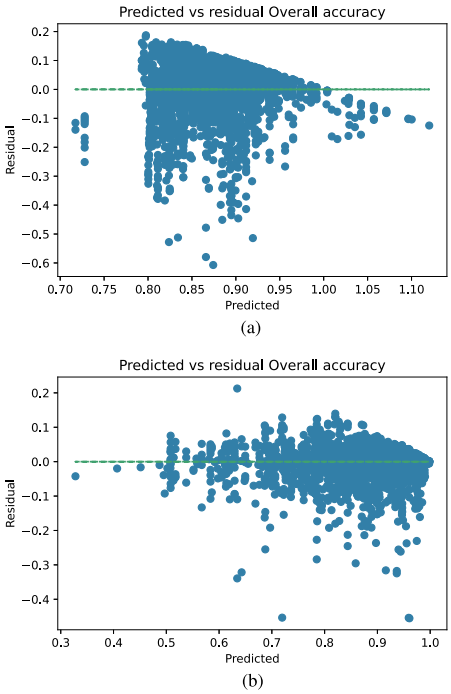


Fig. 16. Residual results for the OA based on (a) OrdLS linear regression and (b) EnsLV nonlinear regression.

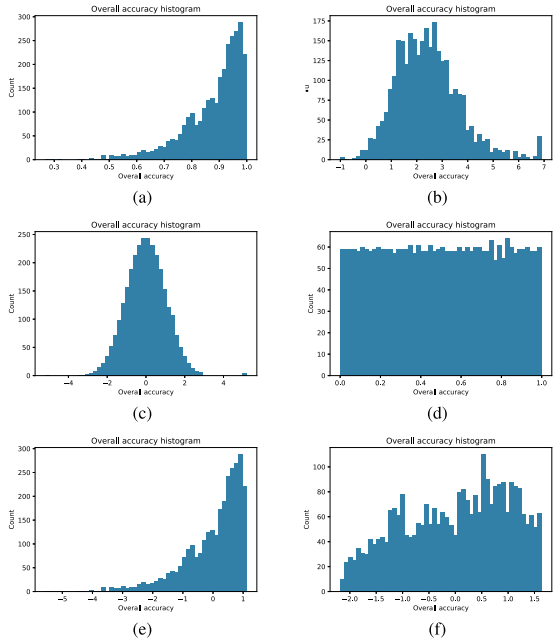


Fig. 17. Histogram of overall accuracy with (a) no transformation, (b) logit transformation, (c) Gaussian transformation, (d) uniform transformation, (e) standard transformation, and (f) power transformation.

predictions and the observations' total variance. The assumption is that the difference between these variances equals the variance of the residual error. Therefore, the R^2 score is always between 0 and 1. This does not hold for the nonlinear methods; hence, R^2 is not reported for the nonlinear methods. The t-statistic is the ratio between the coefficient and the standard error

$$t_i = \frac{\hat{\beta}_i \sqrt{n}}{\sigma(\hat{\beta}_i)} \tag{9}$$

where $\sigma(\hat{\beta}_i)$ is the standard error of $\hat{\beta}_i$. The t-statistic has an associated p-value that can be defined as the probability that it is more extreme than the t-statistic under the null hypothesis H_0 (no effect), i.e., $p_i = Pr(|t| \geq T|H_0)$.

D. Linear Regression Results

Table IV shows the results of the linear regression analysis. It can be seen that the MAE is 0.074 and R^2 is 0.183 for the OA. The low R^2 along with low p-values for several predictors indicates that the regression model can somewhat capture the overall trend in the data but has high variance in prediction errors. For more precise predictions, a higher R^2 is desired and to achieve that, few techniques are possible, such as finding another predictor, transforming the data, or trying a nonlinear model.

Fig. 16(a) shows the residual between the prediction values of the linear regression model used in the regression analysis for OA and the observation values. There is considerable heteroskedasticity in the residual. The OrdLS finds a linear combination of input features and is optimized based on the squared distance between the prediction values of model and all the samples. One of the assumptions of a good linear model is that the error, or residual, between the model and samples has near constant variance across all prediction values. This is known as homoskedasticity. Conversely, heteroskedasticity is changing variance in residuals across prediction values. As shown in Fig. 16(a), OrdLS has similar residual variance on the lower end of the predictions. As the prediction values are moved up, the residual variance increases on the low side and decreases on the high side. Another assumption of linear regression is that the prediction values are unbounded in each direction. However, the linear regression is used based on the OAs that have lower and upper bounds at 0% and 100%. This restriction violates the second assumption. Fig. 16(a) shows the phenomenon, where residuals on the high side trend toward zero. We can also see that some prediction values are outside the bounds. What this shows is that a linear regression model is insufficient to deal with the data we have collected and to get a more accurate prediction model we must try other models.

It is desired to increase the precision in our prediction and stay within the bounds of the dependent variables. In order to do these five transformations, i.e., logit, Gaussian, uniform, standard, and power transformations, on the dependent variable are done and run through a number of linear and nonlinear models. Fig. 17 shows the distribution of OA with and without transformations by the logit, Gaussian, uniform, standard, and power transformations.

The models are tuned using a cross-validated grid search and all experiments are performed through a 5-fold cross-validation. The experimental results are reported in Table V. It can be seen that applying transformations do not significantly improve results and there is no remarkable difference between different models.

E. Nonlinear Regression Results

Table VI shows the experimental results for the nonlinear prediction methods for the dependent variable OA with or

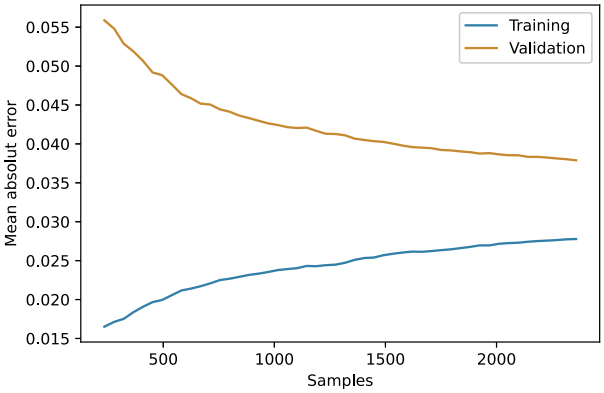


Fig. 18. Learning curve of the EnsLV.

without transformations by logit, Gaussian, uniform standard, and power. Fig. 17 shows the distribution of OA with or without transformations by the logit, Gaussian, uniform, standard, and power transformations. All experiments are conducted through a 5-fold cross-validation. It can be seen that comparing with the linear prediction methods in Table V, the nonlinear prediction methods yield much lower MAEs. This illustrates that nonlinear prediction methods are more suitable for predicting classification performance based on the metadata than linear prediction methods. From Table VI, it can also be observed that the proposed EnsLV method is better than all the other compared nonlinear prediction methods for the dependent variable OA with or without transformations by logit, Gaussian, uniform standard, and power. The EnsLV method with logit transformation applied achieves the best results. Fig. 18 shows that the learning curve of the EnsLV starts to converge at 2000 training samples. The bias displayed between the training and validation curve demonstrates the noise in the data.

Fig. 16(b) shows the residual between the prediction values of the EnsLV nonlinear method used in the regression analysis for OA and the observation values. From Fig. 16(b), it can be seen that there is significant improvement in heteroskedasticity compared to the OrdLS and all predicted values fall within the bounds of the dependent variable. This further demonstrates that the proposed EnsLV method can yield good prediction results.

VI. Conclusion

In this article, we proposed a novel method to analyze to what extent HSIs can be classified by using only metadata. The metadata containing 17 types of different features, such as the classifier categories, the type of selecting training samples, and the number of classes in the dataset, was collected from the 100 HSI classification papers and was used for estimating the performance of classifiers before conducting any experiments. This can provide some guidelines for researchers to identify optimal classifiers and propose classification methods. For example, before doing the real experiments, researchers can estimate the HSI classification results of their classification approaches using our proposed prediction method based on specific parameters in Table I. Then, the results can be used for checking the effectiveness of their classification approaches. In the experiments, some linear and nonlinear prediction methods were trained and tested and analyzed based on the metadata. The experimental results indicate that the proposed EnsLV yields best results. Our future work will focus on sampling more data which will add more results about different datasets and methods and improve our prediction. At the same time, we will improve the precision of our prediction method by considering more elaborate features for the metadata.

References

[1] A. Goetz and G. Vane, "Imaging spectrometry for earth remote sensing," *Science*, vol. 228, no. 4704, pp. 1147–1153, 1985.
[2] D. Haboudane, J. R. Miller, E. Pattey, P. J. Zarco-Tejada, and I. B. Strachan, "Hyperspectral vegetation indices and novel algorithms for predicting green LAI of crop canopies: Modeling and validation in the context of precision agriculture," *Remote Sens. Environ.*, vol. 90, no. 3, pp. 337–352, 2004.
[3] F. A. Kruse, "Mapping surface mineralogy using imaging spectrometry," *Geomorphology*, vol. 137, no. 1, pp. 41–56, 2012.
[4] H. Ren and C. Chang, "Automatic spectral target recognition in hyperspectral imagery," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 39, no. 4, pp. 1232–1249, Oct. 2003.
[5] F. Melgani and L. Bruzzone, "Classification of hyperspectral remote sensing images with support vector machines," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 8, pp. 1778–1790, Aug. 2004.
[6] P. Ghamisi, J. Benediktsson, and M. Ulfarsson, "Spectral-spatial classification of hyperspectral images based on hidden Markov random fields," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 5, pp. 2565–2574, May 2014.
[7] Y. Chen, Z. Z. H. Lin, X. Zhao, G. Wang, and Y. Gu, "Deep learning-based classification of hyperspectral data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2094–2107, Jun. 2014.
[8] P. Ghamisi, J. Plaza, Y. Chen, J. Li, and A. Plaza, "Advanced supervised spectral classifiers for hyperspectral images: A review," *IEEE Geosci. Remote Sens. Mag.*, vol. 5, no. 1, pp. 8–32, Mar. 2017.
[9] P. Ghamisi *et al.*, "New frontiers in spectral-spatial hyperspectral image classification: The latest advances based on mathematical morphology, Markov random fields, segmentation, sparse representation, and deep learning," *IEEE Geosci. Remote Sens. Mag.*, vol. 6, no. 3, pp. 10–43, Sep. 2018.
[10] R. Khatami, G. Mountrakis, and S. V. Stehman, "A meta-analysis of remote sensing research on supervised pixel-based land-cover image classification processes: General guidelines for practitioners and future research," *Remote Sens. Environ.*, vol. 177, pp. 89–100, 2016.
[11] G. G. Wilkinson, "Results and implications of a study of fifteen years of satellite image classification experiments," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 433–440, Mar. 2005.
[12] X. Yang *et al.*, "Hyperspectral image classification with deep learning models," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 9, pp. 5408–5423, Sep. 2018.
[13] S. Li *et al.*, "Deep learning for hyperspectral image classification: An overview," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 9, pp. 6690–6709, Sep. 2019.
[14] M. Fauvel, Y. Tarabalka, J. A. Benediktsson, J. Chanussot, and J. C. Tilton, "Advances in spectral-spatial classification of hyperspectral images," *Proc. IEEE*, vol. 101, no. 3, pp. 652–675, Mar. 2013.
[15] H. Chen, F. Miao, and X. Shen, "Hyperspectral remote sensing image classification with CNN based on quantum genetic-optimized sparse representation," *IEEE Access*, vol. 8, pp. 99900–99909, 2020.
[16] Z. Xue, P. Du, and H. Su, "Kernelized sparse graph-embedded dimensionality reduction for hyperspectral image classification," in *Proc. 6th Workshop Hyperspectral Image Signal Process.: Evol. Remote Sens.*, 2014, pp. 1–4.
[17] M. Feurer *et al.*, "Efficient and robust automated machine learning," *Adv. Neural Inf. Process. Syst.*, vol. 28, pp. 1–9, 2015.
[18] C. Cortes and V. Vapnik, "Support vector networks," *Mach. Learn.*, vol. 20, pp. 273–297, 1995.
[19] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.
[20] R. M. Haralick, S. R. Sternberg, and X. Zhuang, "Image analysis using mathematical morphology," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-9, no. 4, pp. 532–550, Jul. 1987.

[21] M. Coster and J. Chermant, "Image analysis and mathematical morphology for civil engineering materials," *Cement Concrete Composites*, vol. 23, no. 2–3, pp. 133–151, 2001.

[22] 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.

[23] M. Dalla Mura, A. Villa, J. A. Benediktsson, J. Chanussot, and L. Bruzzone, "Classification of hyperspectral images by using extended morphological attribute profiles and independent component analysis," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 3, pp. 542–546, May 2011.

[24] M. Dalla Mura, J. A. Benediktsson, B. Waske, and L. Bruzzone, "Extended profiles with morphological attribute filters for the analysis of hyperspectral data," *Int. J. Remote Sens.*, vol. 31, no. 22, pp. 5975–5991, 2010.

[25] B. Song *et al.*, "Remotely sensed image classification using sparse representations of morphological attribute profiles," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 8, pp. 5122–5136, Aug. 2014.

[26] M. Fauvel, J. Chanussot, and J. A. Benediktsson, "Kernel principal component analysis for the classification of hyperspectral remote sensing data over urban areas," *EURASIP J. Adv. Signal Process.*, vol. 2009, pp. 1–14, 2009.

[27] M. Dalla Mura, J. A. Benediktsson, B. Waske, and L. Bruzzone, "Morphological attribute filters for the analysis of very high resolution remote sensing images," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, vol. 3, 2009, pp. III–97.

[28] M. Fauvel, J. A. Benediktsson, J. Chanussot, and J. R. Sveinsson, "Spectral and spatial classification of hyperspectral data using SVMs and morphological profiles," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 11, pp. 3804–3814, Nov. 2008.

[29] X. Zhang, Y. Sun, and W. Qi, "Hyperspectral image classification based on extended morphological attribute profiles and abundance information," in *Proc. 9th Workshop Hyperspectral Image Signal Process.: Evol. Remote Sens.*, 2018, pp. 1–5.

[30] J. Liao and L. Wang, "Adaptive hyperspectral image classification based on the fusion of manifolds filter and spatial correlation features," *IEEE Access*, vol. 8, pp. 90390–90409, 2020.

[31] B. Song, J. Li, P. Li, and A. Plaza, "Decision fusion based on extended multi-attribute profiles for hyperspectral image classification," in *Proc. 5th Workshop Hyperspectral Image Signal Process.: Evol. Remote Sens.*, 2013, pp. 1–4.

[32] N. Alajlan, Y. Bazi, and R. R. Yager, "Ensemble classification of hyperspectral images based on ordered weighted averaging operators," in *Proc. 5th Workshop Hyperspectral Image Signal Process.: Evol. Remote Sens.*, 2013, pp. 1–4.

[33] X. Huang *et al.*, "Multiple morphological profiles from multicomponent-base images for hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 12, pp. 4653–4669, Dec. 2014.

[34] K. Wang, R. Huang, and Q. Song, "Spectral-spatial hyperspectral image classification using extended multi attribute profiles and guided bilateral filter," in *Proc. Int. Conf. Comput. Sci. Mech. Automat.*, 2015, pp. 235–239.

[35] P. Sidike, C. Chen, V. Asari, Y. Xu, and W. Li, "Classification of hyperspectral image using multiscale spatial texture features," in *Proc. 8th Workshop Hyperspectral Image Signal Process.: Evol. Remote Sens.*, 2016, pp. 1–4.

[36] C. Chen, J. Jiang, B. Zhang, W. Yang, and J. Guo, "Hyperspectral image classification using gradient local auto-correlations," in *Proc. 3rd IAPR Asian Conf. Pattern Recognit.*, 2015, pp. 454–458.

[37] A. Plaza, P. Martinez, J. Plaza, and R. Perez, "Dimensionality reduction and classification of hyperspectral image data using sequences of extended morphological transformations," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 466–479, Mar. 2005.

[38] B. Waske, S. Linden, J. A. Benediktsson, A. Rabe, and P. Hostert, "Impact of different morphological profiles on the classification accuracy of urban hyperspectral data," in *Proc. 1st Workshop Hyperspectral Image Signal Process.: Evol. Remote Sens.*, 2009, pp. 1–4.

[39] A. Plaza and J. Plaza, "Parallel morphological classification of hyperspectral imagery using extended opening and closing by reconstruction operations," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, vol. 1, 2008, pp. I-58–I-61.

[40] M. Jouni, M. D. Mura, and P. Comon, "Hyperspectral image classification using tensor cp decomposition," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2019, pp. 1164–1167.

[41] J. Li, J. M. Bioucas-Dias, and A. Plaza, "Spectral-spatial classification of hyperspectral data using loopy belief propagation and active learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 2, pp. 844–856, Feb. 2013.

[42] P. R. Marpu, M. Pedergnana, M. Dalla Mura, J. A. Benediktsson, and L. Bruzzone, "Automatic generation of standard deviation attribute profiles for spectral-spatial classification of remote sensing data," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 2, pp. 293–297, Mar. 2013.

[43] P. Quesada-Barriuso, F. Argüello, and D. B. Heras, "Spectral-spatial classification of hyperspectral images using wavelets and extended morphological profiles," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 4, pp. 1177–1185, Apr. 2014.

[44] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-6, no. 6, pp. 721–741, Nov. 1984.

[45] Y. Zhong, X. Lin, and L. Zhang, "A support vector conditional random fields classifier with a Mahalanobis distance boundary constraint for high spatial resolution remote sensing imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 4, pp. 1314–1330, Apr. 2014.

[46] C. I. Chang, *Hyperspectral Imaging: Techniques for Spectral Detection and Classification*, vol. 1.Secaucus, NJ, USA: Springer, 2003.

[47] H. Ghanbari, S. Homayouni, A. Safari, and P. Ghamisi, "Gaussian mixture model and Markov random fields for hyperspectral image classification," *Eur. J. Remote Sens.*, vol. 51, no. 1, pp. 889–900, 2018.

[48] Y. Hu, S. T. Monteiro, and E. Saber, "Comparing inference methods for conditional random fields for hyperspectral image classification," in *Proc. 7th Workshop Hyperspectral Image Signal Process.: Evol. Remote Sens.*, 2015, pp. 1–4.

[49] X. Cao *et al.*, "Hyperspectral image classification with Markov random fields and a convolutional neural network," *IEEE Trans. Image Process.*, vol. 27, no. 5, pp. 2354–2367, May 2018.

[50] Y. Tarabalka, M. Fauvel, J. Chanussot, and J. A. Benediktsson, "SVM and MRF-based method for accurate classification of hyperspectral images," *IEEE Geosci. Remote Sens. Lett.*, vol. 7, no. 4, pp. 736–740, Oct. 2010.

[51] L. Sun, Z. Wu, J. Liu, L. Xiao, and Z. Wei, "Supervised spectral-spatial hyperspectral image classification with weighted Markov random fields," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 3, pp. 1490–1503, Mar. 2015.

[52] G. Moser and S. B. Serpico, "Combining support vector machines and Markov random fields in an integrated framework for contextual image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 5, pp. 2734–2752, May 2013.

[53] B. Zhang, S. Li, X. Jia, L. Gao, and M. Peng, "Adaptive Markov random field approach for classification of hyperspectral imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 5, pp. 973–977, Sep. 2011.

[54] U. B. Gewali and S. T. Monteiro, "Spectral angle based unary energy functions for spatial-spectral hyperspectral classification using Markov random fields," in *Proc. 8th Workshop Hyperspectral Image Signal Process.: Evol. Remote Sens.*, 2016, pp. 1–6.

[55] B. C. Kuo, C. H. Chuang, C. S. Huang, and C. C. Hung, "A nonparametric contextual classification based on Markov random fields," in *Proc. 1st Workshop Hyperspectral Image Signal Process.: Evol. Remote Sens.*, 2009, pp. 1–4.

[56] Y. Xu, Z. Wu, and Z. Wei, "Markov random field with homogeneous areas priors for hyperspectral image classification," in *Proc. IEEE Geosci. Remote Sens. Symp.*, 2014, pp. 3426–3429.

[57] P. Ghamisi, J. A. Benediktsson, and M. O. Ulfarsson, "The spectral-spatial classification of hyperspectral images based on hidden Markov random field and its expectation-maximization," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2013, pp. 1107–1110.

[58] A. Ma and A. M. Filippi, "Hyperspectral image classification via object-oriented segmentation-based sequential feature extraction and recurrent neural network," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2020, pp. 72–75.

[59] Q. Jackson and D. Landgrebe, "Adaptive Bayesian contextual classification based on Markov random fields," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 11, pp. 2454–2463, Nov. 2002.

[60] Y. Yuan, J. Lin, and Q. Wang, "Hyperspectral image classification via multitask joint sparse representation and stepwise MRF optimization," *IEEE Trans. Cybern.*, vol. 46, no. 12, pp. 2966–2977, Dec. 2016.

[61] P. Ghamisi, M. S. Couceiro, F. M. L. Martins, and J. A. Benediktsson, "Multilevel image segmentation based on fractional-order darwinian particle swarm optimization," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 5, pp. 2382–2394, May 2014.

[62] B. Zhao, L. Gao, W. Liao, and B. Zhang, "A new kernel method for hyperspectral image feature extraction," *Geo-Spatial Inf. Sci.*, vol. 20, no. 4, pp. 309–318, 2017.

[63] Y. Tarabalka, J. A. Benediktsson, and J. Chanussot, "Spectral-spatial classification of hyperspectral imagery based on partitional clustering techniques," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 8, pp. 2973–2987, Aug. 2009.

[64] J. Li, J. M. Bioucas-Dias, and A. Plaza, "Spectral-spatial hyperspectral image segmentation using subspace multinomial logistic regression and Markov random fields," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 3, pp. 809–823, Mar. 2012.

[65] X. Kang, S. Li, and J. A. Benediktsson, "Spectral-spatial hyperspectral image classification with edge-preserving filtering," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 5, pp. 2666–2677, May 2014.

[66] Y. Tarabalka, J. A. Benediktsson, J. Chanussot, and J. C. Tilton, "Multiple spectral-spatial classification approach for hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 11, pp. 4122–4132, Nov. 2010.

[67] D. Akbari, S. Homayouni, A. Safari, S. Khazai, and H. Torabzadeh, "An improved marker selection method for hyperspectral image segmentation and classification," in *Proc. 6th Workshop Hyperspectral Image Signal Process.: Evol. Remote Sens.*, 2014, pp. 1–4.

[68] Y. Tarabalka and J. C. Tilton, "Spectral-spatial classification of hyperspectral images using hierarchical optimization," in *Proc. 3rd Workshop Hyperspectral Image Signal Process.: Evol. Remote Sens.*, 2011, pp. 1–4.

[69] L. Fang, S. Li, W. Duan, J. Ren, and J. A. Benediktsson, "Classification of hyperspectral images by exploiting spectral-spatial information of superpixel via multiple kernels," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 12, pp. 6663–6674, Dec. 2015.

[70] Y. Chen, N. M. Nasrabadi, and T. D. Tran, "Hyperspectral image classification using dictionary-based sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 10, pp. 3973–3985, Oct. 2011.

[71] L. Gao *et al.*, "Spectral superresolution of multispectral imagery with joint sparse and low-rank learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 3, pp. 2269–2280, Mar. 2021.

[72] L. Fang, S. Li, X. Kang, and J. A. Benediktsson, "Spectral-spatial hyperspectral image classification via multiscale adaptive sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 12, pp. 7738–7749, Dec. 2014.

[73] X. Zhang, P. Weng, J. Feng, E. Zhang, and B. Hou, "Spatial-spectral classification based on group sparse coding for hyperspectral image," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2013, pp. 1745–1748.

[74] M. Cui and S. Prasad, "Multiscale sparse representation classification for robust hyperspectral image analysis," in *Proc. IEEE Glob. Conf. Signal Inf. Process.*, 2013, pp. 969–972.

[75] A. Sumarsono and Q. Du, "Hyperspectral image classification with low-rank subspace and sparse representation," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2015, pp. 2864–2867.

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

[77] Y. Qian, M. Ye, and J. Zhou, "Hyperspectral image classification based on structured sparse logistic regression and three-dimensional wavelet texture features," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 4, pp. 2276–2291, Apr. 2013.

[78] H. Zhang, J. Li, Y. Huang, and L. Zhang, "A nonlocal weighted joint sparse representation classification method for hyperspectral imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2056–2065, Jun. 2014.

[79] H. Yu, X. Zhang, M. Song, J. Hu, and L. Gao, "Superpixel-level constraint representation for hyperspectral imagery classification," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2020, pp. 56–59.

[80] L. Fang, S. Li, X. Kang, and J. A. Benediktsson, "Spectral-spatial classification of hyperspectral images with a superpixel-based discriminative sparse model," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 8, pp. 4186–4201, Aug. 2015.

[81] L. Fang, G. Liu, S. Li, P. Ghamisi, and J. A. Benediktsson, "Hyperspectral image classification with squeeze multibias network," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 3, pp. 1291–1301, Mar. 2019.

[82] G. Zhao, G. Liu, L. Fang, B. Tu, and P. Ghamisi, "Multiple convolutional layers fusion framework for hyperspectral image classification," *Neurocomputing*, vol. 339, pp. 149–160 2019.

[83] Y. Chen *et al.*, "Automatic design of convolutional neural network for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 9, pp. 7048–7066, Sep. 2019.

[84] S. K. Roy, G. Krishna, S. R. Dubey, and B. B. Chaudhuri, "Hybridsn: Exploring 3-D-2-D CNN feature hierarchy for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 2, pp. 277–281, Feb. 2020.

[85] A. Santara *et al.*, "BASS net: Band-adaptive spectral-spatial feature learning neural network for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 9, pp. 5293–5301, Sep. 2017.

[86] R. S. W. Chu, H. Ng, X. Wang, and W. Luk, "Convolution based spectral partitioning architecture for hyperspectral image classification," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2019, pp. 3962–3965.

[87] M. E. Paoletti *et al.*, "Deep pyramidal residual networks for spectral-spatial hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 2, pp. 740–754, Feb. 2019.

[88] W. Wang, S. Dou, Z. Jiang, and L. Sun, "A fast dense spectral-spatial convolution network framework for hyperspectral images classification," *Remote Sens.*, vol. 10, pp. 1068–1086, 2018.

[89] K. Makantasis, K. Karantzalos, A. Doulamis, and N. Doulamis, "Deep supervised learning for hyperspectral data classification through convolutional neural networks," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2015, pp. 4959–4962.

[90] M. Zhao, C. Yu, M. Song, and C. Chang, "A semantic feature extraction method for hyperspectral image classification based on hashing learning," in *Proc. 9th Workshop Hyperspectral Image Signal Process.: Evol. Remote Sens.*, 2018, pp. 1–5.

[91] S. K. Roy, S. R. Dubey, S. Chatterjee, and B. B. Chaudhuri, "Fusenet: Fused squeeze-and-excitation network for spectral-spatial hyperspectral image classification," *IET Image Process.*, vol. 14, no. 8, pp. 1653–1661, 2020.

[92] X. Jiang *et al.*, "Spectral-spatial hyperspectral image classification using dual-channel capsule networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 18, no. 6, pp. 1094–1098, Jun. 2021.

[93] L. Mou, P. Ghamisi, and X. X. Zhu, "Deep recurrent neural networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 7, pp. 3639–3655, Jul. 2017.

[94] J. Zheng, Y. Feng, C. Bai, and J. Zhang, "Hyperspectral image classification using mixed convolutions and covariance pooling," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 1, pp. 522–534, Jan. 2021.

[95] M. Zhu, L. Jiao, F. Liu, S. Yang, and J. Wang, "Residual spectral-spatial attention network for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 1, pp. 449–462, Jan. 2021.

[96] H. Huang, C. Pu, Y. Li, and Y. Duan, "Adaptive residual convolutional neural network for hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, no. 1, pp. 2520–2531, 2020.

[97] S. Wan *et al.*, "Hyperspectral image classification with context-aware dynamic graph convolutional network," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 1, pp. 597–612, Jan. 2021.

[98] Z. Zhong, J. Li, L. Ma, H. Jiang, and H. Zhao, "Deep residual networks for hyperspectral image classification," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2017, pp. 1824–1827.

[99] Y. Chen, H. Jiang, C. Li, X. Jia, and P. Ghamisi, "Deep feature extraction and classification of hyperspectral images based on convolutional neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 10, pp. 6232–6251, Oct. 2016.

[100] I. Bidari *et al.*, "Hyperspectral imagery classification using deep learning," in *Proc. 4th World Conf. Smart Trends Syst., Secur. Sustainability*, 2020, pp. 672–676.

[101] Z. Li, H. Huang, and C. Pu, "Deep manifold learning network for hyperspectral image classification," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2020, pp. 2021–2024.

[102] Y. Li, L. Zhang, W. Wei, and Y. Zhang, "Deep self-supervised learning for few-shot hyperspectral image classification," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2020, pp. 501–504.

[103] C. Xing, L. Ma, and X. Yang, "Stacked denoise autoencoder based feature extraction and classification for hyperspectral images," *J. Sensors*, vol. 2016, pp. 1–10.

[104] S. Hao, W. Wang, Y. Ye, T. Nie, and L. Bruzzone, "Two-stream deep architecture for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 4, pp. 2349–2361, Apr. 2018.

[105] H. Patel and K. P. Upla, "A shallow network for hyperspectral image classification using an autoencoder with convolutional neural network," *Multimedia Tools Appl.*, vol. 81, no. 1, pp. 695–714, 2022.

[106] H. Zhang, Y. Li, Y. Zhang, and Q. Shen, "Spectral-spatial classification of hyperspectral imagery using a dual-channel convolutional neural network," *Remote Sens. Lett.*, vol. 8, no. 5, pp. 438–447, 2017.

[107] J. Wang, X. Song, L. Sun, W. Huang, and J. Wang, "A novel cubic convolutional neural network for hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, no. 1, pp. 4133–4148, 2020. .
[108] Y. Chen *et al.*, "Hyperspectral images classification with Gabor filtering and convolutional neural network," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 12, pp. 2355–2359, Dec. 2017.
[109] H. Liang and Q. Li, "Hyperspectral imagery classification using sparse representations of convolutional neural network features," *Remote Sens.*, vol. 8, no. 2, pp. 99–114, 2016.
[110] R. J. Williams and D. Zipser, "A learning algorithm for continually running fully recurrent neural networks," *Neural Comput.*, vol. 1, no. 2, pp. 270–280, 1989.
[111] P. Rodriguez, J. Wiles, and J. L. Elman, "A recurrent neural network that learns to count," *Connection Sci.*, vol. 11, no. 1, pp. 5–40, 1999.
[112] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.
[113] A. Graves, "Generating sequences with recurrent neural networks," 2013, *arXiv:1308.0850*.
[114] M. E. Paoletti, J. M. Haut, J. Plaza, and A. Plaza, "Scalable recurrent neural network for hyperspectral image classification," *J. Supercomputing*, vol. 76, no. 11, pp. 8866–8882, 2020.
[115] A. Ma, A. M. Filippi, Z. Wang, and Z. Yin, "Hyperspectral image classification using similarity measurements-based deep recurrent neural networks," *Remote Sens.*, vol. 11, no. 2, pp. 194–212, 2019.
[116] L. Zhu, Y. Chen, P. Ghamisi, and J. A. Benediktsson, "Generative adversarial networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 9, pp. 5046–5063, Sep. 2018.
[117] T. Dam, N. Swami, S. G. Anavatti, and H. A. Abbass, "Multi-fake evolutionary generative adversarial networks for imbalance hyperspectral image classification," 2021, *arXiv:2111.04019*.

Bin Zhao (Member, IEEE) received the B.S. degree in remote sensing science and technology and the B.S. degree in international economics and trade from the Shandong Agricultural University, Taian, China, in 2014, the M.S. degree in electronics and communication engineering from the Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, University of Chinese Academy of Sciences, Beijing, China, in 2017, and the Ph.D. degree in electrical and computer engineering from the University of Iceland, Reykjavik, Iceland, in 2021.

He is currently a Research Associate with the University of Iceland. His research interests include hyperspectral image denoising, superresolution, feature extraction, and classification.

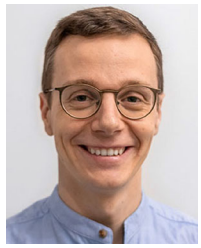
Haukur Isfeld Ragnarsson (Student Member, IEEE) received the B.S. degree in music composition from the Iceland Academy of the Arts, Reykjavík, Iceland, in 2009, and the B.S. degree in electrical and computer engineering and the M.S. degree in electronics and communication engineering from the University of Iceland, Reykjavík, Iceland, in 2015 and 2021, respectively.

His research interests include hyperspectral image classification.

Magnus O. Ulfarsson (Senior Member, IEEE) received the B.S. and M.S. degrees from the University of Iceland, Reykjavík, Iceland, in 2002, and the Ph.D. degree from the University of Michigan, Ann Arbor, MI, USA, in 2007, all in electrical and computer engineering.

In 2007, he joined the University of Iceland, where he is currently a Professor and the Chair of the Faculty of Electrical and Computer Engineering. Since 2013, he has been with deCODE Genetics, Reykjavík, Iceland. His research interests include statistical signal processing, image processing, machine learning, remote sensing, medical imaging, and genomics.

Dr. Ulfarsson has been an Associate Editor for IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING since 2018.



Gabriele Cavallaro (Member, IEEE) received the B.Sc. and M.Sc. degrees in telecommunications engineering from the University of Trento, Trento, Italy, in 2011 and 2013, respectively, and the Ph.D. degree in electrical and computer engineering from the University of Iceland, Reykjavík, Iceland, in 2016.

He is currently the Head of the AI and ML for Remote Sensing Simulation and Data Lab, Jülich Supercomputing Centre, Wilhelm-Johnen-Strasse, Juelich, Germany. Since 2019, he has been giving lectures on scalable machine learning for remote sensing Big Data at the Institute of Geodesy and Geoinformation, University of Bonn, Bonn, Germany. His research interests cover remote sensing data processing with parallel machine learning algorithms that scale on high-performance and distributed systems.

Dr. Cavallaro was the recipient of the IEEE GRSS Third Prize in the Student Paper Contest of the *2015 IEEE International Geoscience and Remote Sensing Symposium* (Milan, Italy). He is the Chair of the High-Performance and Disruptive Computing in Remote Sensing Working Group of the IEEE Geoscience and Remote Sensing Society (GRSS) Earth Science Informatics Technical Committee.

Jón Atli Benediktsson (Fellow, IEEE) received the Cand.Sci. degree in electrical engineering from the University of Iceland, Reykjavík, Iceland, in 1984, and the M.S.E.E. and Ph.D. degrees in electrical engineering from Purdue University, West Lafayette, IN, USA, in 1987 and 1990, respectively.

Since 2015, he has been the President and a Rector with the University of Iceland. From 2009 to 2015, he was the Pro-Rector of Science and Academic Affairs and a Professor of electrical and computer engineering with the University of Iceland. He is a Cofounder of the biomedical start-up company Oxymap, Reykjavík, Iceland. His research interests are in remote sensing, biomedical analysis of signals, pattern recognition, image processing, and signal processing. He has authored or coauthored extensively in those fields.

Dr. Benediktsson was a member of the 2014 IEEE Fellow Committee. He is a member of Academia Europea, the Association of Chartered Engineers in Iceland (VFI), Societas Scientiarum Islandica, and Tau Beta Pi. He is a fellow of the Society of Photo-Optical Instrumentation Engineers (SPIE). He was the 2011–2012 President of the IEEE Geoscience and Remote Sensing Society (GRSS). He has been on the GRSS AdCom since 2000. He was the Chairperson of the Steering Committee of *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* (JSTARS) from 2007 to 2010. He was a recipient of the Icelandic Research Council's Outstanding Young Researcher Award in 1997 and the IEEE Third Millennium Medal in 2000. In 2021, he was honored as a recipient of the Order of the Falcon from the President of Iceland. He was a corecipient of the University of Iceland's Technology Innovation Award in 2004, the 2012 IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING Paper Award, the IEEE GRSS Highest Impact Paper Award in 2013, and the *International Journal of Image and Data Fusion* Best Paper Award in 2016 and 2018. He was the recipient of the Stevan J. Kristof Award from Purdue University in 1991 as an outstanding graduate student in remote sensing, the Yearly Research Award from the Engineering Research Institute of the University of Iceland in 2006, the Outstanding Service Award from the IEEE GRSS in 2007, the IEEE/VFI Electrical Engineer of the Year Award in 2013, the OECE Award from the School of ECE, Purdue University, in 2016, the IEEE GRSS David Landgrebe Award in 2018, and the IEEE GRSS Education Award in 2020. He serves on the International Editorial Board of the *International Journal of Image and Data Fusion* and the Editorial Board of *Remote Sensing*. He was the Editor-in-Chief for the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING (TGRS) from 2003 to 2008. He has been serving as an Associate Editor for IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING since 1999, the *IEEE Geoscience and Remote Sensing Letters* since 2003, and *IEEE Access* since 2013. He is currently a Senior Editor for the *Proceedings of the IEEE*. He was a Highly Cited Researcher (Clarivate Analysis) from 2018 to 2021.