Hyperspectral Image Classification—Traditional to Deep Models: A Survey for Future Prospects

Muhammad Ahmad, Sidrah Shabbir, Swalpa Kumar Roy, *Student Member, IEEE*, Danfeng Hong, *Senior Member, IEEE*, Xin Wu, *Member, IEEE*, Jing Yao, Adil Mehmood Khan, *Member, IEEE*, Manuel Mazzara, Salvatore Distefano, and Jocelyn Chanussot, *Fellow, IEEE*

***Abstract*—Hyperspectral imaging (HSI) has been extensively utilized in many real-life applications because it benefits from the detailed spectral information contained in each pixel. Notably, the complex characteristics, i.e., the nonlinear relation among the captured spectral information and the corresponding object of HSI data, make accurate classification challenging for traditional methods. In the last few years, deep learning (DL) has been substantiated as a powerful feature extractor that effectively addresses the nonlinear problems that appeared in a number of computer vision tasks. This prompts the deployment of DL for HSI classification (HSIC) which revealed good performance. This survey enlists a systematic overview of DL for HSIC and compared state-of-the-art strategies of the said topic. Primarily, we will encapsulate the main challenges of TML for HSIC and then we will acquaint the superiority of DL to address these problems. This article breaks down the state-of-the-art DL frameworks into spectral-features, spatial-features, and together spatial–spectral features to systematically analyze the achievements (future research directions as well) of these frameworks for HSIC. Moreover, we will consider the fact that DL requires a large number of labeled training examples whereas acquiring such a number for HSIC is challenging in terms of time and cost. Therefore, this survey discusses some strategies to improve the generalization performance of DL strategies which can provide some future guidelines.**

***Index Terms*—Deep learning (DL), feature learning, hyperspectral image classification (HSIC), hyperspectral imaging (HSI), spectral–spatial information.**

Manuscript received October 24, 2021; revised November 19, 2021; accepted November 30, 2021. Date of publication December 9, 2021; date of current version January 20, 2022. This work was supported in part by the National Natural Science Foundation of China under Grant 42030111, Grant 41722108, and Grant 62101045, in part by China Postdoctoral Science Foundation under Grant 2021M690385, in part by MIAI@Grenoble Alpes under Grant ANR-19-P3IA-0003, in part by AXA Research Fund, in part by the Analytical Center for the Government of the Russian Federation under Grant 70-2021-00143 dd. 01.11.2021 and Grant IGK 000000D730321P5Q0002. *(Corresponding author: Danfeng Hong.)*

Muhammad Ahmad is with the Department of Computer Science, National University of Computer and Emerging Sciences, Islamabad, Chiniot-Faisalabad Campus, Chiniot 35400, Pakistan, and also with the Dipartimento di Matematica e Informatica—MIFT, University of Messina, 98121 Messina, Italy (e-mail: mahmad00@gmail.com).

Sidrah Shabbir is with the Department of Computer Engineering, Khwaja Fareed University of Engineering and Information Technology (KFUEIT), Rahim yar khan 64200, Pakistan (e-mail: sidrah.shabbir@gmail.com).

Swalpa Kumar Roy is with the Department of Computer Science and Engineering, Jalpaiguri Government Engineering College, West Bengal 735102, India (e-mail: swalpa@cse.jgec.ac.in).

Danfeng Hong and Jing Yao are with the Key Laboratory of Digital Earth Science, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China (e-mail: hongdf@aircas.ac.cn; jasonyao92@gmail.com).

Xin Wu is with the School of Information and Electronics, Beijing Institute of Technology, Beijing 100081, China (e-mail: 040251522wuxin@163.com).

Adil Mehmood Khan is with the Institute of Data Science and Artificial Intelligence, Innopolis University, 420500 Innopolis, Russia (e-mail: a.khan@innopolis.ru).

Manuel Mazzara is with the Institute of Software Development and Engineering, Innopolis University, 420500 Innopolis, Russia (e-mail: m.mazzara@innopolis.ru).

Salvatore Distefano is with the Dipartimento di Matematica e Informatica—MIFT, University of Messina, 98121 Messina, Italy (e-mail: sdistefano@unime.it).

Jocelyn Chanussot is with the CNRS, Grenoble INP, GIPSA-Lab, Universite Grenoble Alpes, 38400 Grenoble, France (e-mail: jocelyn.chanussot@grenoble-inp.fr).

Digital Object Identifier 10.1109/JSTARS.2021.3133021

I. Introduction

HYPERSPECTRAL imaging (HSI) is concerned with the extraction of meaningful information based on the radiance acquired by the sensor at short or long distances without substantial contact with the object of interest [1]. HSI provides detailed spectral information by sampling the reflective portion of the electromagnetic spectrum covering a wide range of 0.4–2.4 m (i.e., visible 0.4–0.7 m to short wave infrared 0.7–2.4 m) region in hundreds of narrow and contiguous spectral bands. HSI can also explore the (light) emission properties of objects in the range of mid to long infrared regions [2].

Despite the detailed information, it brings several challenges since traditional analysis techniques for monochromatic, RGB, and multispectral images cannot be directly exploited to extract meaningful information from hyperspectral ones due to several reasons, e.g., HSI exhibits the unique statistical and geometrical properties of high-dimensional spectral/spatial data, i.e., the volume of a hypercube and hypersphere concentrates on corners and outside shells, respectively.

HSI has been adopted in several real-world applications, including but not limited to the atmosphere, environmental, urban, agriculture, geological and mineral exploration, coastal zone, marine, forestry (i.e., track forest health), water quality and surface contamination, inland waters, and wetlands, snow and ice, biological, medical contexts, and food processing [3]–[8]. There are also several military applications in camouflage, landmine detection, and littoral zone mapping. Furthermore, HSI has been used in space, air, and underwater vehicles to acquire detailed spectral information for a wide range of uses [9]–[12].

Infield collection and spectral library indexing of ground truth signatures for any of the said applications are critical for many

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. Various real-world applications of HSI.

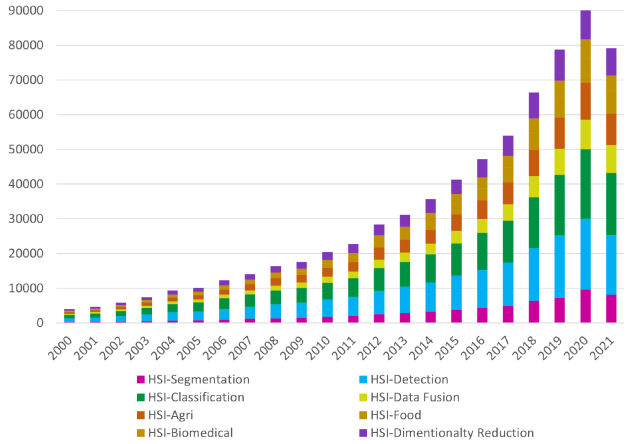


Fig. 2. Various HSI-related articles published per year till September 25, 2021 [Source: Google Scholar accessed on September 25, 2021 and the results (including patents and citations) were sorted by relevance].

reasons. For instance, the spectral information of vegetation is prejudiced by a wide range of environmental situations that make it challenging to satisfactorily represent variability without the collection of site-specific field spectra. But the real potential of HSI is mostly untapped since it allows it to go deeper than surface features, considering that usually, each feature has a different spectrum band. HSI, indeed, can capture more than 200 spectral bands which help practitioners to discriminate objects that were not possible before. A few HSI application examples are shown in Fig. 1, but several other domains (e.g., smart city, Industry 4.0, Intelligent Transportation Systems) can greatly benefit from such an approach.

Considering the aforementioned limitations, HSI analysis is categorized into the following main streams: dimensionality reduction (DR) [13]–[17], spectral unmixing [18]–[26], object/change detection [27]–[33] classification [34]–[36], feature learning for classification [37]–[41], restoration and denoising [42], [43], and resolution enhancement [44], [45]. Fig. 2 shows an exponentially growing trend in the literature published per year for HSI analysis-related tasks and applications.

In this article, we specifically focus on HSI data classification (HSIC), which has achieved a phenomenal interest of the research community due to its broad applications in the areas of land use and land cover [46]–[50], environment monitoring and natural hazards detection [51], [52], vegetation mapping [53], [54], and urban planning. HSIC methodologies exploit machine learning algorithms to perform the classification task [55], [56]. These methods are outlined in various comprehensive reviews published during/in the last decade [34], [57]–[65]. Nevertheless, continuous advancements in the field of machine learning provide improved methods from time to time. Deep learning (DL) models is one of such revolutionary advancements in machine learning that improved HSIC accuracy [66]–[68].

This article aims to give an overview of the widely used DL-based techniques to perform HSIC. Specifically, we will first summarize the main challenges of HSIC which cannot be effectively overcome by traditional machine learning (TML), and later we will enlist the advantages of DL to handle the aforementioned issues. At a later stage, we will provide a framework to categorize the corresponding works among the following.

1) Spectral and spatial feature learning, individually.
2) Spectral–spatial feature learning to systematically review the achievements in DL-based HSIC.
3) Future research stems to improve the generalization performance and robustness of DL models while considering the limited availability of reliable training samples.

The remainder of this article is structured as follows. Section II introduces the task of HSI classification (HSIC) and briefly discusses the HSIC paradigm shift from TML (conventional) to DL models, describing HSI data characteristics along with the advantages and limitations of DL that are faced while working with HSI. In Sections III and IV, we give an overview of different forms of HSI representations and basic machine learning strategies, respectively. Section V describes a few commonly used types of layers and reviews recent developments (specifically from 2017 onward) of some intensively utilized DL frameworks for HSIC. Sections VI–IX presents the state-of-the-art developments of convolutional neural networks (CNNs), graph CNN (GCNN), autoencoders (AEs), deep belief networks (DBNs), and recurrent neural networks (RNNs), respectively. In Section X, we briefly discussed various strategies to overcome the low generalization performance of HSIC due to the limited availability of training data. Section XI presents the experimental results and discussion on results obtained using different DL strategies. Section XII concludes the article with a few future research directions related to joint exploitation of spectral-spatial features of HSI, limited training data, and computational complexity.

II. Hyperspectral Image Classification (Background and Challenges)

A. Traditional to DL Models

The main task of HSIC is to assign a unique label to each pixel vector of HSI cube based on its spectral or spectral–spatial properties. Mathematically, an HSI cube can be represented as $\mathbf{X} = [x_1, x_2, x_3, \ldots, x_B]^T \in \mathcal{R}^{B\times(N\times M)}$, where B represents the total number of spectral bands consisting of $(N \times M)$ samples per band belonging to $\mathbf{Y}$ classes, where

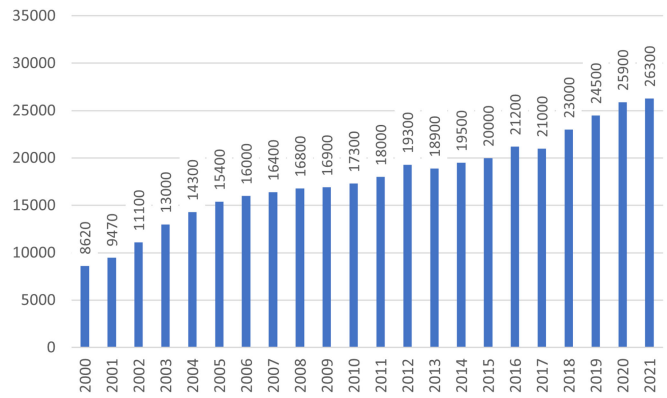


Fig. 3. Remote sensing/hyperspectral image classification-related articles published per year till September 25, 2021 [Source: Google Scholar accessed on September 25, 2021 and the results (including patents and citations) were sorted by relevance].

$x_i = [x_{1,i},\ x_{2,i},\ x_{3,i}, \ldots, x_{B,i}]^T$ is the ith sample in the HSI cube with class label $y_i \in \mathcal{R}^Y$. The classification problem can be considered as an optimization one, in which a mapping function $f_c(.)$ takes the input data $\mathbf{X}$ and, after applying some transformations over it, obtains the corresponding label $\mathbf{Y}$, to reduce the gap between obtained output and the actual one [69].

$$Y = f_c(X, \theta) \tag{1}$$

where θ is a certain adjustable parameter that may be required to apply transformations on input data $\mathbf{X}$ such that $f_c \colon X \rightarrow Y$.

In the literature, substantial work has been done on HSIC and there is a growing trend in the development of such techniques as shown in Fig. 3. Most HSIC frameworks seemed to be influenced by the methodologies used in the computer vision domain [70]. TML-based HSIC approaches use handcrafted features to train the classifier. These methods generally rely on utilizing engineering skills and domain expertise to design several human-engineered features, for instance, shape, texture, color, shape, spectral, and spatial details. All these features are basic characteristics of an image and carry effective information for image classification. Commonly used handcrafted feature extraction (FE) and classification methods include texture descriptors such as local binary patterns (LBPs) [71], histogram of oriented gradients (HOG) [72], global image scale-invariant transform/global invariant scalable transform (GIST) [73], pyramid HOGs, scale-invariant feature transform (SIFT) [74], random forests (RF) [75], kernel-based support vector machine (SVM) [76], K-nearest neighbors (KNN), and extreme learning machine (ELM).

Color histograms are simple and effective handcrafted features used for an image classification task. They are easy to compute and invariant to small changes in images, i.e., translation and rotation. The major drawback of a color histogram is that it does not provide spatial contextual information, hence it becomes difficult to distinguish between objects of the same color but different distribution. Moreover, color histograms are sensitive to variance in illumination. HOG features represent the histogram of edge orientations of spatial subregions. It can effectively extract the edge and local shape details and has been utilized in various remote sensing-related works [46], [77]–[79].

SIFT is a broadly used robust feature descriptor applied to image classification tasks [80]–[83]. The advantage of the SIFT descriptor is that it is invariant to the changes in image scale, rotation, illumination, and noise. SIFT is used to extract local features that describe a specific point in the image. The disadvantage of SIFT is that it is mathematically complex, which increases its computational cost. GIST represents the global description of important aspects of an image, that is, the scales and orientations (gradient information) of various subregions of an image. GIST builds a spatial envelope in terms of different statistical properties like roughness, openness, and ruggedness, etc [84]. Texture descriptors such as LBPs are used for remote sensing image analysis [71], [85]. LBPs are used to describe the texture around each pixel by choosing pixels from the square neighborhood and gray level values of all neighborhood pixels are thresholded with respect to the central pixel.

The color histograms, GIST, and texture descriptors are global features that represent certain statistical characteristics of an image like color, texture [86], [87], and spatial structure [73]. HOG and SIFT are local features that describe geometrical information. Usually they are used to construct bag-of-visual-words (BoVW) models [48], [52], [83], [88]–[93] and HOG feature-based models [46], [94]. Some popular feature encoding or pooling strategies to enhance the performance of BoVW are Fisher vector coding [71], [95], [96], spatial pyramid matching [97], and probabilistic topic model [93], [98]–[100]. A single feature is insufficient to represent the whole image information, and hence a combination of these features is used for image classification [47], [88], [98], [100]–[107].

Handcrafted features can effectively represent the various attributes of an image, and hence working well with the data being analyzed. However, these features may be insubstantial in the case of real data; therefore, it is difficult to fine-tune between robustness and discriminability as a set of optimal features considerably vary between different data. Furthermore, human involvement in designing the features considerably affects the classification process, as it requires a high level of domain expertise to design handcrafted features.

To mitigate the limitations of handcrafted feature designing, a deep feature learning strategy was proposed by Hinton and Salakhutdinov [108]. DL-based methods can automatically learn the features from data in a hierarchical manner, to construct a model with growing semantic layers until a suitable representation is achieved. Such models have shown great potential for feature representation in remote sensing image classification [109], [110].

DL architectures can learn the behavior of any data without any prior knowledge regarding the statistical distribution of the input data [111] and can extract both linear and nonlinear features of input data without any pre-specified information. Such systems are capable of handling HSI data in both spectral and spatial domains individually, and also in a coupled fashion. DL systems possess a flexible architecture in terms of types of layers and their depth and are adaptive to various machine learning

strategies like supervised, semi-supervised, and unsupervised techniques.

B. Hyperspectral Data Characteristics and DL Challenges

Despite the above-discussed DL potentials, there are still some challenges that need to be considered while applying DL to HSI data. Most of these challenges are related to the characteristics of HSI data, i.e., hundreds of contiguous and narrow spectral channels with very high spectral resolution and low spatial resolution throughout the electromagnetic spectrum coupled with limited availability of training data. Although the pixels with rich spectral information are useful for classification purposes, however, the computation of such data takes a lot of time and resources.

Furthermore, processing such high-dimensional data is a somewhat complex task due to an increased number of parameters. This is known as the curse of dimensionality which considerably influences the classification performance, especially in the case of supervised learning [112]. Since the size of training data is not adequate/insufficient and/or not reliable (i.e., the training samples may not provide any new information to the model or may have similar patterns/structures) to properly train the classifier, it may lead the model to overfit. This is known as the Hughes phenomena [113] which occur when labeled training data is significantly smaller than the number of spectral bands present in the data. Lack of labeled HSI data is a major issue in HSIC as labeling of HSI is a time-consuming and expensive task because it usually requires human experts or investigation of real-time scenarios.

In addition to high dimensionality, HSIC suffers from various other artifacts like high intraclass variability due to unconfined variations in reflectance values caused by several environmental interferers and degradation of data caused by instrumental noise while capturing the data [114]. Furthermore, the addition of redundant bands due to HSI instruments affects the computational complexity of the model. Spectral mixing is another challenge related to the spatial resolution of HSI. HSI pixels with low to average spatial resolution cover vast spatial regions on the surface of earth, leading to mixed spectral signatures which result in high interclass similarity in border regions. As a result, it becomes difficult to identify the materials based on their spectral reflectance values [115]. The following are some main challenges that come across when DL is applied to HSIC.

- *Complex Training Process:* Training of deep neural network (DNN) and optimization by tuning parameters is an NP-complete problem, where the convergence of the optimization process is not guaranteed [116]–[118]. Therefore, it is assumed that training of DNN is very difficult [111], especially in the case of HSI when a large number of parameters need to be adjusted/tuned. However, the convergence task becomes somehow easier due to the advancement of various optimization techniques for deep CNNs. Among stochastic gradient descent [119] and its momentum version [120], RMSProp [121], Adam [122], AdamW [123], diffGrad [124], RAdam [125], gradient centralization (GC) [126], and AngularGrad [127], respectively, are the successful CNN optimization techniques and widely used in any classification problems.
- *Limited Availability of Training Data:* As discussed above, supervised DNN requires a considerably large amount of training data; otherwise, their tendency to overfit increases significantly [128], which leads to the Hughes phenomena. The high-dimensional characteristic of HSI coupled with a small amount of labeled training data makes the DNNs ineffective for HSIC as it demands a lot of adjustments during the training phase [69].
- *Model's Interpretability:* The training procedure of DNNs is difficult to interpret and understand. The black box kind of nature is considered as a potential weakness of DNNs and may affect the design decisions of the optimization process. However, a lot of work has been done to interpret the model's internal dynamics.
- *High Computational Burden:* One of the main challenges of DNN is dealing with a big amount of data that involves increased memory bandwidth, high computational cost, and storage consumption [129]. However, advanced processing techniques like parallel and distributed architectures [130], [131] and high-performance computing (HPC) [115] make it possible for DNNs to process large amounts of data.
- *Training Accuracy Degradation:* It is assumed that deeper networks extract more rich features from data [132]; however, this is not true for all systems to achieve higher accuracy by simply adding more layers, because by increasing the network's depth, the problem of exploding or vanishing gradient becomes more prominent [133] and affects the convergence of the model [132].

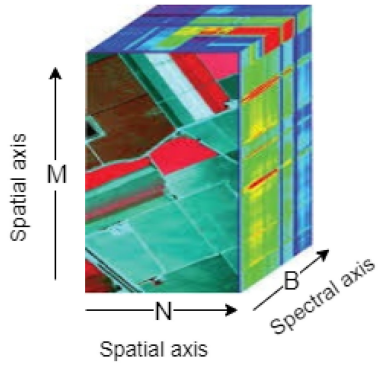


Fig. 4. Hyperspectral cube.

III. HSI Representation

Hyperspectral data is represented in the form of a 3-D hypercube, $X \in \mathcal{R}^{B\times(N\times M)}$, which contains 1-D spectral and 2-D spatial details of a sample, where B represents the total number of spectral bands and N and M are spatial components, i.e., width and height, respectively. The HSI cube is shown in Fig. 4.

A. Spectral Representation

In such representations, each pixel vector is isolated from other pixels and processed based on spectral signatures only, which means the pixel is represented only in spectral space $x_i \in \mathcal{R}^B$, where B can either be the actual number of spectral channels or just relevant spectral bands extracted after some DR method. Usually, instead of using original spectral bands, a low dimensional representation of HSI is preferred for data processing in order to avoid redundancy and achieve better class separability, without considerable loss of useful information.

DR approaches for spectral HSI representation can either be supervised or unsupervised. Unsupervised techniques transform the high-dimensional HSI into a low-dimensional space without using the class label information; for example, principal component analysis (PCA) and locally linear embedding [134]. On the other hand, supervised DR methods utilize labeled samples to learn the data distribution. i.e., to keep data points of the same classes near to each other and separate the data points of different classes. For instance, linear discriminant analysis (LDA), local Fisher discriminant analysis (LFDA) [135], local discriminant embedding [136], and nonparametric weighted FE [137]. LDA and LDFA provide better class separability by maximizing the interclass distance of data points and minimizing the intraclass distance. However, due to the spectral mixing effect, in which the same material may appear with different spectra or different materials may have the same spectral signatures, it becomes difficult to differentiate among different classes based on the spectral reflectance values alone.

B. Spatial Representation

To deal with the limitations of spectral representation, another approach is to exploit the spatial information of the pixels, in which pixels in each band are represented in the form of a matrix, $x_i \in \mathcal{R}^{N \times M}$. Due to high spatial correlation, neighboring pixels have higher probabilities to belong to the same class. Therefore, in the case of spatial representation, neighboring pixels' information is also considered and the neighborhood of a pixel can be determined using kernel- or pixel-centric window [138]. Some common methods to extract spatial information from HSI cube are morphological profiles (MPs), texture features [like Gabor filters, gray-level co-occurrence matrix (GLCM), and LBP, etc.] and DNN-based methods. Morphological profiles are capable of extracting geometrical characteristics. Few extensions of MPs include extended morphological profiles (EMPs) [139], multiple-structure-element morphological profiles [140], and invariant attribute profiles [141].

The texture of the image provides useful spatial contextual information of HSI. For instance, a Gabor filter, a texture analysis technique, can efficiently obtain textural information at various scales and orientations. Similarly, LBP can provide rotation-invariant spatial texture representation. The GLCM can effectively determine the spatial variability of HSI by exploiting the relative positions of neighborhood pixels. The DNNs can also extract spatial information of HSI by considering the pixel as an image patch instead of representing it as a spectral vector. The spatial information contained in HSI can also be extracted by combining various of the aforediscussed methods. For instance, [142] combined Gabor filter and differential morphological profiles [143] to extract local spatial sequential (LSS) features for a RNN-based HSIC framework.

C. Spectral–Spatial Representation

This representation jointly exploits both spectral and spatial information of data. In such approaches, a pixel vector is processed based on spectral features while considering spatial-contextual information. The strategies that simultaneously use both spectral and spatial representations of HSI either concatenate the spatial details with spectral vector [62], [144] or process the 3-D HSI cube to preserve the actual structure and contextual information [145].

In the literature, all these HSI representations are widely exploited for HSIC. Most of the DNNs for pixel-wise classification utilized the spectral representation of HSIs [146], [147]. However, to mitigate the limitations of spectral representation, many efforts have been made to incorporate the spatial information [148], [149]. Recently, joint exploitation of both spectral and spatial features has gained much popularity and led to improved classification accuracy [13], [67], [150]–[153]. These HSI feature exploitation approaches, for HSIC, are further discussed in Section IV.

IV. Learning Strategies

DL models can adopt various learning strategies that can be broadly categorized into the following.

A. Supervised Learning

In a supervised learning approach, the model is trained based on the labeled training data, which means training data is composed of a set of inputs and their corresponding outputs or class labels. During the training phase, the model iteratively updates its parameters in order to predict the desired outputs accurately. In the testing phase, the model is tested against the new input/test data in order to validate its ability to predict the correct labels. If trained sufficiently, the model can predict the labels of new input data. However, supervised learning of DNNs requires a lot of labeled training data to fine-tune the model parameter. Therefore, they are best suited to scenarios where plentiful labeled data is available. The details of various supervised learning techniques for DNNs will be explained in the respective sections.

B. Unsupervised Learning

In contrast to the supervised learning approach, unsupervised learning techniques learn from the input data with no explicit labels associated with it. These approaches try to identify the underlying statistical structure of input representations or patterns in the absence of corresponding labels. As there is no ground truth available for the training data so it might be difficult to measure the accuracy of the trained model. However, such learning strategies are useful in the cases where we want to learn the inherent structure of such datasets which have a scarcity of

training data. The PCA is an unsupervised learning technique that can be used to learn a low-dimensional representation of the input. Similarly, k-means clustering is another unsupervised learning method that groups the input data into homogeneous clusters.

C. Semi-Supervised Learning

The semi-supervised learning (SSL) technique is halfway between unsupervised and supervised approaches. Semi-supervised techniques learns from the partially labeled data i.e., a small amount of labeled training data can be utilized to label the rest of the unlabeled data. These techniques effectively utilize all available data instead of just labeled data; therefore, these techniques have gained much popularity among the research community and are being widely used for HSIC [154]–[157]. The details of these methods are briefly described in Section X.

V. Development of DNNs (Types of Layers)

In the following, we review recent developments of some widely used DNN frameworks for HSIC. We specifically surveyed the literature published from 2017 onward. DNNs exhibit a great variety of flexible and configurable models for HSIC that allow the incorporation of several types of layers. Few widely used types of layers are explained in the following.

A layer is the key building block of DNN, and the type of layer has a decisive impact in terms of feature processing. A layer takes the weighted input, processes it through linear or nonlinear transformation, and outputs these values to the next layer. Generally, a layer is uniform, as it has a single activation function. The first layer of the network is known as the input layer and the last layer as an output layer. All other layers in the network, in between the input and output layers, are known as hidden layers. These layers progressively find different features in the input data by performing various transformations. The choice of layer type depends on the task at hand, as some layers perform better for some tasks than others. The most commonly used layers for HSIC are explained below.

A. Fully Connected Layers

A fully connected (FC) layer connects every neuron in the lower layer to every neuron in the upper/next layer. Mostly, they are used as the last few layers of a model usually after convolution/pooling layers. FC takes the output of the previous layer and assigns weights to predict the probabilities for class labels. Due to a large number of connections, a large number of parameters need to be adjusted, which significantly increases the computational overhead. Moreover, due to a large number of parameters, the model becomes more sensitive to overfitting [49]. However, to mitigate the effect of overfitting, a dropout method is introduced in [158].

B. Convolutional Layers

The convolutional (CONV) layer convolves the input data or feature maps from a lower layer with the filters (kernels). The filter contains weights whose dot product is calculated with the subset of input data by moving it across the width, height, and depth of the input region. The output of the filter is known as a feature map. CONV layer provides spatial invariance via a local connectivity approach in which the neuron in the feature map connects to a subset of input from the previous layer rather than connecting to every neuron. This reduces the number of parameters that need to train. To further reduce the number of parameters, the CONV layer uses the mechanism of parameter sharing in which the same weights are used in a particular feature map.

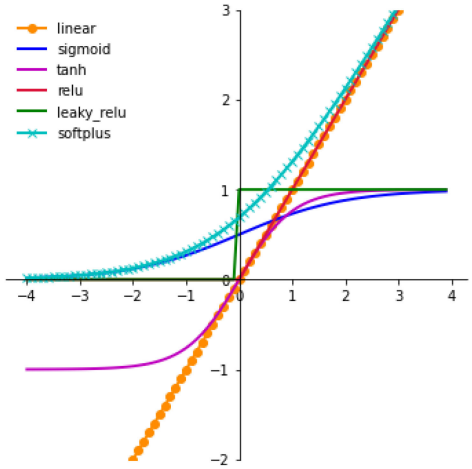


Fig. 5. Graphical representation of various commonly used activation functions.

C. Activation Layers

Activation layers are assumed to be a feature detector stage of DNNs [159]. FC and CONV layers provide linear representations of input data or it can be said that they work similarly to linear regressors and data transformed by these layers is considered to be at the FE stage [69]. Therefore, to learn nonlinear features of data, an activation layer must be used after FC and CONV layers. In the activation layer, feature maps from previous layers go through an activation function to form an activation map. Some commonly used activation functions are sigmoid, hyperbolic tangent (tanh), rectified linear unit (ReLU), LiSHT [160], and softmax. However, in HSI analysis, softmax and ReLU are widely employed activation functions [69]. Fig. 5 presents a graphical representation of a few commonly utilized activation functions.

D. Pooling or Subsampling Layers

The pooling layer, also known as the subsampling or downsampling layer, takes a certain input volume and reduces it to a single value as shown in Fig. 6. This provides invariance to small distortions in the data. The pooling layer helps the model to control overfitting as the size of data and model parameters both are reduced, which also leads to a decrease in the computational time. The commonly used downsampling operations are max-pooling, average-pooling, and sum-pooling. Recently, a pooling technique, wavelet-pooling, is introduced in [161], whose

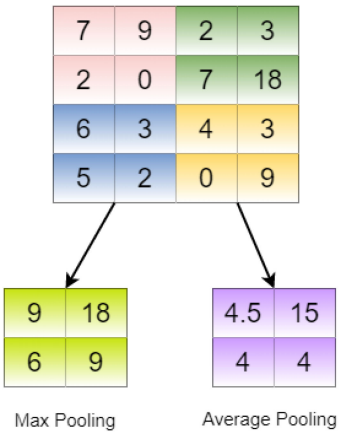


Fig. 6. Max-pooling and average-pooling operations of downsampling/pooling layer.

performance is commensurable to max-pooling and average-pooling. Alternatively, [162] proposed another trend in which the pooling layer is replaced by the CONV layer of increased filter stride.

VI. Convolutional Neural Network

The architecture of the CNN is inspired by the biological visual system presented in [163]. Following the natural visual recognition mechanism proposed by Hubel and Wiesel [163], Neocognitron [164] is regarded as the first hierarchical, position-invariant model for pattern recognition [165], which can be considered as the predecessor of CNN [166]. The architecture of CNN can be divided into two main stages: one is FE network and the other is a classification based on the feature maps extracted in the first stage.

The FE network consists of multiple hierarchically stacked CONV, activation, and pooling layers. The CONV layer extracts the features from input data by convolving a learned kernel with it. On each CONV layer, the kernel is spatially shared with whole input data, which reduces the model's complexity and the network becomes easier to train as the number of parameters that need to be fine-tuned is reduced. Convolved results are then passed through an activation layer which adds nonlinearities in the network to extract nonlinear features of the input. This is achieved by applying a nonlinear function to the convolved results. Afterward, the resolution of the feature map is reduced by applying a pooling operation to achieve shift-invariance. Generally, the pooling layer is added with every CONV layer followed by the activation function.

The classification stage consisting of FC layers and a Softmax operator gives the probability of input pattern belonging to a specific class based on the feature maps extracted at the FE stage. FC layer connects every single neuron in the previous layer to every neuron in the current layer. In Refs. [167] and [168], the authors proposed that the FC layer can be disregarded by using a global average pooling layer. Softmax is commonly used for classification tasks [169], [170]; however, many works have also utilized SVM [171], [172] for this purpose.

In the following, we reviewed three types of CNN architectures for HSIC: i) Spectral CNN, ii) spatial CNN, and iii) spectral–spatial CNN. Fig. 7 illustrates the general architecture of these three frameworks.

A. Spectral CNN Frameworks for HSIC

Spectral CNN models only consider 1-D spectral information ($x_i \in \mathcal{R}^B$) as input, where B could either be the original number of spectral bands or the appropriate number of bands extracted after some DR method. In [173], a CNN structure was proposed to mitigate the overfitting problem and achieved a better generalization capability by utilizing 1×1 convolutional kernels and enhanced dropout rates. Moreover, a global average pooling layer is used in place of a FC layer in order to reduce the network parameters. To reduce high correlation among HSI bands, Gao *et al.* [168] proposed a CNN architecture for HSIC, which fully utilized the spectral information by transforming the 1-D spectral vector to a 2-D feature matrix, and by cascading composite layers consisting of 1×1 and 3×3 CONV layers, the architecture achieved the feature reuse capability. Similar to [173], [168] also utilized the global average pooling layer to lower the network's training parameters and to extract high-dimensional features.

Wu and Prasad [174] presented a hybrid model for HSIC in which the first few CONV layers are employed to extract position-invariant middle-level features and then recurrent layers are used to extract spectral-contextual details. Similarly, Jin *et al.* [146] used a hybrid architecture for classifying healthy and diseased wheat heads. For the input layer, they transform spectral information into a 2-D data structure. In [175], CNN was proved to be more effective as compared to SVM and KNN for the spectral-based identification of rice seed's variety. A similar application of CNN was explored in [147], where various varieties of *Chrysanthemum* were identified using spectral data of the first five PCs of PCA. PCA is a DR method that is widely used in many DL applications to handle/preprocess high-dimensional data. In [176], PCA was utilized to preprocess medical HSI, and then the fusion of CNN kernels with Gabor kernels using dot product is used for classification.

The study [177] analyzed another DR technique dynamic mode decomposition, which converted 3-D HSI data to 2-D and then this data is fed to vectorized CNN for classification. To overcome the noise effect in pixel-wise HSIC, a method of averaged spectra is used in [178], where an averaged spectra of a group of pixels belonging to bacterial colonies is extracted for further analysis.

B. Spatial CNN Frameworks for HSIC

Spatial CNN models only consider spatial information, and to extract the spatial information from HSI data, DR methods are employed on spectral-domain to lower the dimensionality of original HSI data. For instance, Li *et al.* [179] used PCA to extract the first PC with refined spatial information and fed it to a fully CNN framework for classification. Similarly, Haut *et al.* [180] trained a spatial-based 2-D-CNN with one PC. In [181], PCA-whitened input data considering three PCs

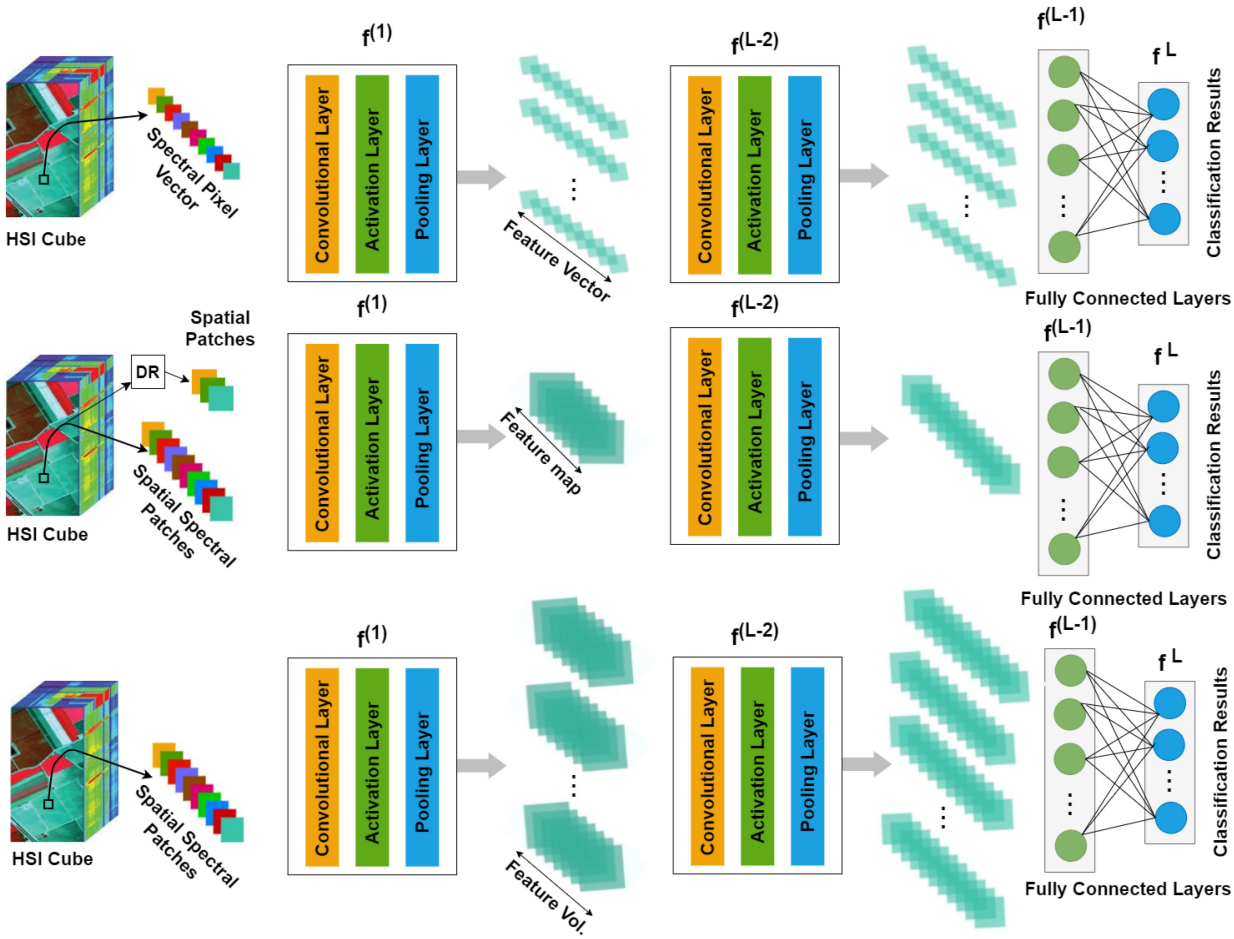


Fig. 7. General architecture of spectral CNN, spatial CNN, and spectral-spatial CNN frameworks for HSIC.

is fed to a random patches network as a 2-D-CNN classification framework. However, the limited training samples with highly similar spectral feature make DL models prone to overfitting. To overcome this, Wang *et al.* [182] proposed a probabilistic neighborhood pooling-based attention network for HSIC.

The method proposed in [183] cropped the patches from 2-D input images (i.e., images from the different spectral bands) to train a 2-D-CNN architecture that learns the data-adaptive kernels by itself. Furthermore, some authors also proposed the utilization of handcrafted features along with spectral-domain reduction. For example, [184] combined the Gabor-filtering technique with 2-D-CNN for HSIC to overcome the overfitting problem due to limited training samples. The Gabor filtering extracts the spatial details, including edges and textures which effectively reduce the overfitting problem. The work [185] proposed a deformable HSIC network based on the concept of deformable sampling locations which can adaptively adjust their size and shape in accordance with HSI's spatial features. Such sampling locations are created by calculating 2-D offsets for every pixel in the input image through regular convolutions by taking into account three PCs. These offsets can cover the locations of similar neighboring pixels possessing similar characteristics. Then, structural information of neighboring pixels is fused to make deformable feature images. Regular convolution employed on these deformable feature images can extract more effective complex structures.

C. Spectral–Spatial CNN Frameworks for HSIC

Spectral–spatial pixel-wise HSIC can be achieved by integrating spatial features into spectral information. For instance, Ran *et al.* [186] presented an improved pixel pair feature (PPF) approach called spatial PPF, which is different from traditional PPFs with respect to two main aspects: One is the selection of pixel pair, that is, only the pixel from the immediate neighborhood of central pixel can be used to make a pair, and second, the label of pixel pair would be as of central pixel. To extract discriminative joint representation, Zhong *et al.* [187] introduced a supervised spectral–spatial residual network that uses a series of 3-D convolutions in the respective spectral and spatial residual blocks. An efficient deep 3-D-CNN framework was proposed in [188] that simultaneously exploits both spectral and spatial information for HSIC.

Similarly, to reflect the variations of spatial contexture in various hyperspectral patches, Li *et al.* [189] implemented an adaptive weight-learning technique instead of assigning fixed weights to incorporate spatial details. Besides this, to make the convolutional kernel more flexible, Roy *et al.* [153] explored

a new architectural design that can adaptively find adjustable receptive feild and then an improved spectral-spatial residual network for joint FE. The discriminative power of the extracted features can be further improved by combining both the max and min convolutional features before the ReLU nonlinearity reported in [190] for the classification task. CNNs failed to exploit rotation equivariance in a natural way; Paoletti *et al.* [191] introduced the translation-equivariant representations of input features which provide extra robustness to the spatial feature locations for HSIC.

The deeper networks may suffer from the issues of overfitting and gradient vanishing problems due to the smaller number of available labeled training samples and to overcome this shortcoming the lightweight CNN's gain good attention in HSIC communities. The paper [192] introduced an end-to-end 3-D lightweight CNN to tackle the limited numbers of training samples for HSIC. To reduce the large gap between the massive trainable parameters and the limited labeled samples, Jia *et al.* [193] proposed to extract the spatial–spectral Schroedinger eigenmaps joint spatial-spectral information, and then further reduced the dimensionality using compression technique. Approximately 90% of trainable weights of the total parameters are used immediately after the flatten operation, i.e., in the FC layer, whereas the remaining only 10% weights are used on the previous convolutional layers of the whole network. To overcome this, the paper [194] introduced a lightweight bag-of-feature learning paradigm into an end-to-end spectral-spatial squeeze-and-excitation residual network for HSIC.

The morphological operations, i.e., erosion and dilation, are powerful nonlinear feature transformations that are widely used to preserve the essential characteristics of shape and structural information of an image. Inspired by these, the paper [195] introduced a new end-to-end morphological CNN (MorphCNN) for HSIC, which utilizes both the spectral and spatial features by concatenating the outputs from spectral and spatial morphological blocks extracted in a dual-path fashion.

The work [189] proposed a two-stage framework for joint spectral–spatial HSIC which can directly extract both spectral and spatial features instead of independently concatenating them. The first stage of the proposed network is composed of a CNN and softmax normalization that adaptively learns the weights for input patches and extracts joint shallow features. These shallow features are then fed to a network of stacked autoencoder (SAE) to obtain deep hierarchical features and final classification is performed with a multinomial logistic regression (MLR) layer. A 3-D-CNN model was introduced in [196] to jointly exploit spectral–spatial features from HSI, and to validate its performance, comparison is performed with spectral-based DBN, SAE, and 2-D-spatial CNN for HSIC. The work [197] introduced a bilinear fusion mechanism over the two branches of squeeze operation based on the global and max-pooling whereas the excitation operation is performed with the fused output of squeeze operation.

The work [198] proposed a deep multiscale spectral-spatial FE approach for HSIC which can learn effective discriminant features from the images with high spatial diversity. The framework utilizes the fully convolutional network (FCN) to extract deep spatial information and, then, these features are fused with spectral information by using a weighted fusion strategy. Finally, pixel-wise classification is performed on these fused features.

In [199], a dual-channel CNN framework was implemented for spectral-spatial HSIC. In the proposed approach, 1-D-CNN is used to hierarchically extract spectral features and 2-D-CNN to extract hierarchical spatial features. These features are then combined together for the final classification task. Furthermore, to overcome the deficiency of training data and to achieve higher classification accuracy, the proposed framework is supported by a data augmentation (DA) technique that can increase the training samples by a factor of 6. In [200], a multiscale 3-D deep CNN is introduced for end-to-end HSIC, which can jointly learn both 1-D spectral and 2-D multiscale spatial features without any preprocessing or postprocessing techniques like PCA, etc. In order to reduce the band redundancy or noise in HSI, [201] explored a novel architecture for HSIC by embedding a band attention module in the traditional CNN framework. The study [202] proposed an HSIC architecture in which PCA-transformed images are used to obtain multiscale cubes for handcrafted FE by utilizing multiscale covariance maps which can simultaneously exploit spectral-spatial details of HSI. These maps are then used to train the traditional CNN model for classification.

The work [203] combined CNN with metric learning-based HSIC framework, which first utilizes CNN to extract deep spatial information using the first three PCs extracted by PCA. Then, in a metric learning-based framework, spectral and spatial features are fused for spectral-spatial feature learning by embedding a metric learning regularization factor for the classifier's training (SVM). Similarly, [204] combines multiscale convolution-based CNN with diversified deep metrics based on determinantal point process (DPP) [205] priors for (1-D spectral, 2-D spectral-spatial, and 3-D spectral-spatial) HSIC. Multiscale filters are used in CNN to obtain multiscale features, and DPP-based diversified metric transformation is performed to increase the interclass variance and decrease intraclass variance, and better HSI representational ability. Final classification maps are obtained by using a softmax classifier.

In a recent work [206], an HSIC framework is proposed to extract multiscale spatial features by constructing a three-channel virtual RGB image from HSI instead of extracting the first three PCs through PCA. The purpose of using a three-channel RGB image is to utilize existing networks trained on natural images to extract spatial features. For multiscale FE, these images are passed to a FCN. These multiscale spatial features are fused and further joined with PCS-extracted spectral features for final classification via SVM.

A two-branch (spectral and spatial) DNN for HSIC was introduced in [207]. The spatial branch consists of a band selection layer and a convolutional and deconvolutional framework with skip architecture to extract spatial information of HSI, and in the spectral branch, a contextual DNN is used to extract spectral features. The paper [208] introduced an adaptive band selection based semi-supervised 3-D-CNN to jointly exploit spectral–spatial features, whereas [209] explored dual-attention based autoencoder–decoder network for

unsupervised hyperspectral band selection and then joint FE for land-cover class prediction. Similarly, in [210], spectral–spatial features are simultaneously exploited in an unsupervised manner using a 3-D convolution AE. The pixel-wise land-use and land-cover classification using traditional CNNs is often suffered by the presence of wrong/noisy labels in the training set and can easily be overfitted to the labeled noises. To overcome this problem of accurate classification, Roy *et al.* [211] proposed a lightweight heterogeneous kernel convolution (HetConv3D) for HSIC with noisy labels by effectively combining both the spectral and spatial kernel features to produce discriminative and invariant feature maps for classification.

A hybrid 3-D-2-D-CNN architecture was presented by Roy *et al.* [212] in which 3-D-CNN is first used to extract joint spectral–spatial features and then 2-D-CNN is further used to obtain more abstract spatial contextual features. The study [213] proposed to use adaptive Markov random field for HSIC. The CNN first extracts joint spectral–spatial features and then a smooth Markov Random Field (MRF) prior is placed on class labels to further refine the spatial details. CNNs are greatly affected by overfitting and vanishing gradient problems, and to overcome this, a separable attention network was introduced by Paoletti *et al.* [214], where the input feature maps are divided into several groups and split along the channel dimension and finally an attention mask encodes global contextual information by combining them. Recently, generalized gradient centralized 3-D convolution (G2C-Conv3D) was introduced in [215] to combine both the intensity-level semantic information and gradient-level detailed information extracted from raw HSIs during the convolutions operation. To boost the performance of accurate land-cover types classification, G2C-Conv3D can be easily plugged into the existing HSIs FE networks.

D. GCN Frameworks for HSIC

GCNs [216] have been garnering increasing attention to researchers in various application fields, owing to their flexible and diversified network architecture that is capable of processing nongrid high-dimensional data. Such properties provide new insight and possibilities in processing hyperspectral data more effectively and efficiently. In detail, GCNs enable the modeling of the relations between data (or samples). Accordingly, this naturally motivates us to use the GCNs to capture the spatial relations of spectral signatures in HSIs. Due to the GCNs' limitations in the graph construction [217], particularly for large graphs (need expensive computational cost), GCNs fail to classify or identify materials in large-scale hyperspectral scenes using normal PCs, which leads to relatively less popularity compared to CNNs in HSIC. For this reason, there have been some tentative researches using the GCNs in the HSIC task.

For example, a second-order GCN was proposed in [218] by modeling spatial–spectral relations on manifolds for HSIC by the attempts to reduce the computational cost on graphs. Wan *et al.* [219] first used superpixel segmentation techniques on HSIs and fed superpixels instead of pixels into GCNs. This enables the network training of GCNs on a large number of pixels in HSIs with the application to the land-cover classification task. Nevertheless, these methods still fail to solve the problem of GCNs essentially. To this end, Hong *et al.* [217] proposed a novel miniGCN. As the name suggests, miniGCN trains the GCNs in a mini-batch fashion, which is the same as CNN. The proposed miniGCN not only reduces the computational cost-effectively but also makes it possible to make a quantitative comparison and fusion with CNNs, further yielding a FuNet for HSIC.

E. Future Directions for CNN-Based HSIC

In the preceding section, we have reviewed the recent developments of CNNs for HSIC. Although CNNs-based HSIC frameworks have achieved great success with respect to classification performance, there are still many aspects that need further investigation. For instance, there is a need to further work on such models that can jointly employ spatial and spectral information for HSIC. Many of the above-surveyed frameworks use DR methods to achieve better spectral–spatial representation but such approaches discard useful spectral information of HSI. Hence, the development of robust HSIC approaches that can preserve spectral information is required. However, the processing of such approaches increases the computational burden, and the training process becomes slower; therefore, parallel processing of such networks using field-programmable gate arrays (FPGAs) and graphical processing units (GPUs) is desired in order to achieve the computationally fast models that can even be suitable for mobile platforms, without the performance degradation.

Moreover, as the CNNs are becoming deeper and deeper, more labeled training data is required for accurate classification, and as discussed before, there is a lack of labeled training data in HSI. In order to overcome this issue, more research is required to integrate the CNN with unsupervised or semi-supervised approaches. Furthermore, we should pay more attention to the generalization ability of CNNs, particularly for the input data format (not only limiting to the grid data). GCNs might be a good solution to combine with CNNs together to develop a more general CNN-based new framework. Using this, we expect to be able to further break the performance bottleneck, yielding more efficient HSIC.

VII. Autoencoders

AE is a popular symmetrical neural network for HSIC due to its unsupervised feature learning capability. AE itself does not perform a classification task, instead it gives a compressed feature representation of high-dimensional HSI data. AE consists of an input layer, one hidden or encoding layer, one reconstruction or decoding layer, and an output layer as shown in Fig. 8. AE is trained on input data in such a manner to encode it into a latent representation that is able to reconstruct the input. To learn a compressed feature representation of input data, AE tries to reduce the reconstruction error that is minimizing the difference between the input and the output.

The SAE is built by stacking multiple layers of AEs in such a way that the output of one layer is served as an input of the subsequent layer. Denoising autoencoder (DAE) is a variant of AE that has a similar structure as AE except for the input data.

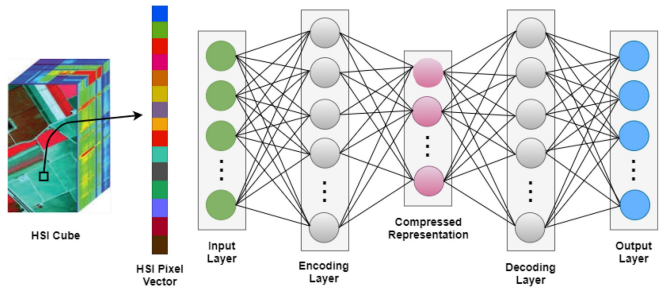


Fig. 8. A general autoencoder architecture.

In DAE, the input is corrupted by adding noise to it; however, the output is the original input signal without noise. Therefore, DAE, different from AE, can recover original input from a noisy input signal.

To learn high-level representation from data, the work [220] proposed a combination of multilayer AEs with maximum noise fraction which reduces the spectral dimensionality of HSI, while a softmax logistic regression classifier is employed for HSIC. The study reported in [221] combined multimanifold learning framework proposed by [222] with counteractive autoencoder [223] for improved unsupervised HSIC. The work [224] jointly exploited spectral–spatial features of HSI through an unsupervised feature extracting framework composed of recursive autoencoders network. It extracts the features from the neighborhood of the target pixel and weights are assigned based on the spectral similarity between target and neighboring pixels. A two-stream DNN with a class-specific fusion scheme was introduced in [225], which learns the fusion weights adaptively. One stream composed of stacked DAE is used to extract spectral features and the second stream is implemented to extract spatial information using CNN, while final classification is performed by fusing the class prediction scores obtained from the classification results of both streams.

Another work proposed a hybrid architecture for multifeature-based spectral–spatial HSIC, which utilizes PCA for DR, guided filters [226] to obtain spatial information, and sparse AE for high-level FE. The framework proposed in [227] exploited both spectral and spatial information for HSIC by adopting batch-based training of AEs and features are generated by fusing spectral and spatial information via a mean pooling scheme. Another work [228] developed a spectral–spatial HSIC framework by extracting appropriate spatial resolution of HSI and utilization of stacked sparse AE for high-level FE followed by RF for the final classification task.

Similarly, [229] also used stacked sparse AE for various types of representation that is spectral–spatial and multifractal features along with other higher order statistical representations. A combination of SAE and ELM was proposed in [230] for HSIC, which segments the features of the training set and transform them via SAE; after transformation, feature subsets are rearranged according to the original order of the training set and fed to ELM-based classifiers, while Q-statistics is used for final classification result. This processing of feature subsets helps to improve variance among base classifiers [230]. Similarly, in a recent work, Ahmad *et al.* [231] implemented a computationally efficient multilayer ELM-based AE, which learns the features in three folds, as proposed in [39] for HSIC.

To overcome the issue of high intraclass variability and high interclass similarity in HSI, Zhou *et al.* [232] developed an SAE-based HSIC, which can learn compact and discriminative features by imposing a local fisher discriminant regularization. Similarly, in the latest work [233], a k-sparse denoising AE is spliced with spectral-restricted spatial features that overcome the high intraclass variability of spatial features for HSIC. The study [234] proposed an HSIC architecture that first makes the spectral segments of HSI based on mutual information measure to reduce the computation time during FE via SAE, while spatial information is incorporated by using EMPs, and SVM/RF is used for final classification. Recently, [235] used SAE for the classification of an oil slick on the sea surface by jointly exploiting spectral–spatial features of HSI.

A. Future Directions for AE-Based HSIC

In the above section, we have surveyed the recent developments of AEs-based techniques for HSIC. Although such frameworks provide powerful predictive performance and show good generalization capabilities, more sophisticated work is still desired. Many of the discussed approaches do not fully exploit abundant spatial information so further techniques need to be developed that can fully employ joint spatial and spectral information for HSIC. Moreover, the issue of high intraclass variability and high inter-class similarity in HSI also hinders the classification performance. Many of the above-reviewed works have addressed this issue but further research to overcome this aforesaid issue is required. One direction could be further exploring approaches like pretraining, cotraining, and adaptive neural networks, etc. for AE-based HSIC frameworks.

VIII. Deep Belief Network

DBN [236] is a hierarchical deep DNN that learns the features from input in an unsupervised, layer-by-layer approach. The layers in DBN are built using restricted Boltzmann machine (RBM) composed of a two-layer architecture in which visible units are connected to hidden units [237], as shown in Fig. 9.

A detailed overview of RBM can be found in [237]. To extract more comprehensive features from input data, the hidden unit of one RBM can be fed to the visible units of other RBM. This type of layer-by-layer architecture builds a DBN, which is trained greedily and can capture deep features from HSI. The architecture of three-layer DBN is shown in Fig. 10.

In literature, several works implemented DBN for HSIC. For instance, Ayhan and Kwan [238] used DBN for land-cover classification by combining spectral–spatial information and making a comparison with some other classification approaches. The usual learning process of DBN involves two steps: One is unsupervised pretraining with unlabeled samples and the second is supervised fine-tuning with the help of labeled samples. However, this training process may result in two problems: First, multiple hidden units may tend to respond similarly [239] due to coadaptation [240] and second is linked with the sparsity and

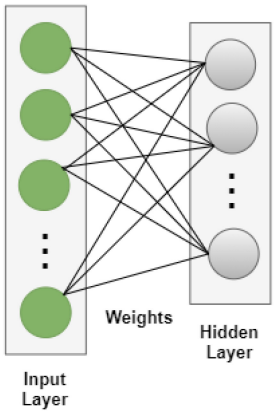


Fig. 9. Basic architecture of RBM.

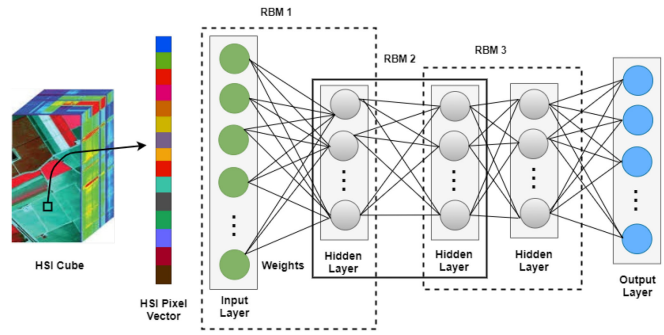


Fig. 10. A three-layer DBN architecture.

selectivity of activated neurons i.e., some neurons may always be dead or always responding [241]. To mitigate these two problems, Zhong *et al.* [242] introduced a diversified DBN model through regularizing the pretraining and fine-tuning process by imposing a diversity prior to enhancing the DBN's classification accuracy for HSI.

To extract efficient texture features for the HSIC, the work [243] proposed a DBN-based texture feature enhancement framework that combines band grouping and sample band selection approach with a guided filter to enhance the texture features, which are then learned by a DBN model, and final classification results are obtained by a softmax classifier. The work [244] implemented a parallel layers framework consisting of Gaussian–Bernoulli RBM which extracts high-level, local invariant, and nonlinear features from HSI, and a logistic regression layer is used for classification.

To improve the classification accuracy, some works are considered to jointly exploit the spectral and spatial information contained in HSI. For instance, Li *et al.* [245] introduced a DBN framework with the logistics regression layer and verified that the joint exploitation of spectral–spatial features leads to improved classification accuracy. Similarly, Sellami and Farah [246] proposed a spectral–spatial graph-based RBM method for HSIC, which constructs the spectral–spatial graph through joint similarity measurement based on spectral and spatial details, then an RBM is trained to extract useful joint spectral–spatial features from HSI, and finally, these features are passed to a DBN and logistic regression layer for classification.

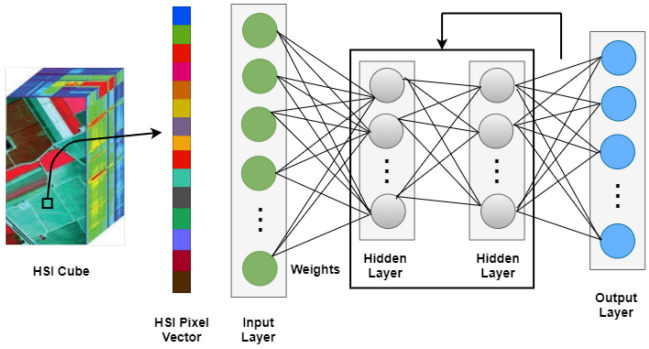


Fig. 11. RNN architecture.

A. Future Directions for DBN-Based HSIC

In the preceding section, we have reviewed the latest developments of DBN-based HSIC frameworks. We have observed that relative to other DNNs, very few works have utilized the DBNs for HSIC. Therefore, there is a need to further explore the DBN-based robust techniques that can jointly employ spatial and spectral features for HSIC. In addition, another research direction can be the regularization of the pretraining and fine-tuning processes of DBN to efficiently overcome the issue of dead or potentially overtolerant (always responding) neurons.

IX. Recurrent Neural Network

The architecture of the RNN, shown in Fig. 11, comprises loop connections, where the node activation of the next step depends on the previous step [247]. Therefore, RNNs are capable of learning temporal sequences. RNN models process the spectral information of HSI data as time sequence considering the spectral bands as time steps [248]. There are three basic models of RNN: a) Vanilla, b) long–short-term memory (LSTM), and c) gated recurrent unit (GRU).

Vanilla is the simplest RNN model and leads to information degradation while processing high-dimensional data. LSTM models composed of two states overcome this issue by controlling the information flow through three gates: input, forget, and output gates. It learns the relevant information over time by discarding the extraneous information. However, the gate-controlling strategy makes the LSTM a considerably complex approach. GRU variant of LSTM enjoys the simplicity of the Vanilla model and provides high performance similar to LSTM. GRU is a simpler version of LSTM, which modifies the input and forget gate as an update (z_t) and reset (r_t) gate and removes the output gate. A comparison of LSTM and GRU's internal architecture is presented in Fig. 12.

The work [70] proposed an RNN-based HSIC framework with a novel activation function (parametric rectified tanh) and GRU, which utilizes the sequential property of HSI to determine the class labels. In [142], a LSS method-based RNN framework was introduced, which first extracts low-level features from HSI by

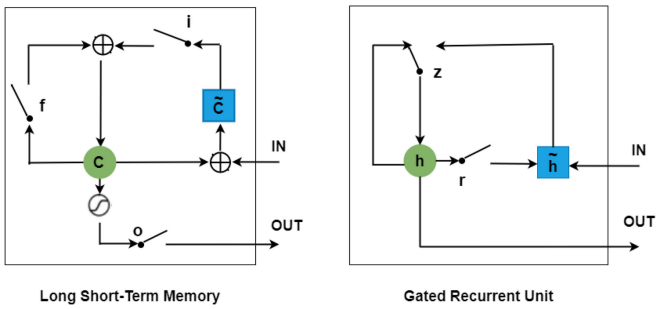


Fig. 12. Internal architecture of LSTM and GRU.

using Gabor filter and differential morphological profiles [143] and then fuses these features to obtain LSS features from the proposed method; these LSS features are further passed to an RNN model to extract high-level features, while a softmax layer is used for final classification.

Keeping in view the usefulness of spatial information to achieve improved classification accuracies, the work [249] proposed a spectral–spatial LSTM-based network that learns spectral and spatial features of HSI by utilizing two separate LSTM-followed softmax layer for classification, while a decision fusion strategy is implemented to get joint spectral–spatial classification results. Similarly, Sharma *et al.* [250] proposed a patch-based RNN with LSTM cells that incorporate multitemporal and multispectral information along with spatial characteristics for land-cover classification.

In literature, several works have proposed CNN-based hybrid RNN architectures (CRNN) for HSIC. For instance, Wu and Prasad [174] implemented a convolutional RNN in which the first few CONV layers are employed to extract position invariant middle-level features, and then recurrent layers are used to extract spectral–contextual details for HSIC. Similarly, Wu and Prasad [251] utilized such a model for semi-supervised HSIC by using pseudo labels. The study [252] suggested an HSIC framework in which CNN is used to extract spatial features from HSI, and then these features are passed to a GRU-based fusion network that performs feature-level and decision-level fusion.

Similarly, Luo, *et al.* [253] exploited both spectral and spatial information contained in HSI by combining CNN with parallel GRU-based RNN, which simplifies the training of GRU and improves performance. Bidirectional convolutional LSTM (CLSTM) was proposed in [152] to jointly exploit spectral–spatial feature of HSI for classification. Shi and Pun [254] combined multiscale local spectral–spatial features extracted by 3-D-CNN with a hierarchical RNN, which learns the spatial dependencies of local spectral–spatial features at multiple scales. Recurrent 2-D-CNN and recurrent 3-D-CNN for HSIC were proposed in [255], and, along with an interesting comparison of these frameworks with their corresponding 2-D and 3-D-CNN models, validates the superiority of recurrent CNN. The work [256] integrated CNN with CLSTM in which a 3-D-CNN model is used to capture low-level spectral-spatial features and CLSTM recurrently analyzes this low-level spectral–spatial information. Recently, Hang *et al.* [70] introduced a cascade RNN for HSIC, which consists of two layers of GRU-based RNN; the first layer is used to reduce the redundant spectral bands and the second layer is used to learn the features from HSI; furthermore, a few convolutional layers are employed to incorporate the rich spatial information contained in HSI.

A. Future Directions for RNN-Based HSIC

In the above section, we have surveyed the recent developments of AEs-based techniques for HSIC. Although RNN-based HSIC frameworks have attracted considerable attention to the remote sensing community and achieved great success for classification performance, there are still many aspects that need further investigation; for instance, the construction of sequential input data for RNN. Most of the surveyed methods considered HSI pixel as a sequential point that is the pixel from each spectral band that forms a data sequence. However, this increases the length of RNN's input sequence considerably large, which can lead to an overfitting issue.

Moreover, processing such large data sequences increases the computational time and the learning process becomes slower. Therefore, the use of parallel processing tools needs to be further investigated to achieve good generalization performance of RNN-based HSIC. In addition, approaches like a grouping of spectral bands to decrease the data sequence length and utilization of the entire spectral signature to better discriminate between various classes can further be explored to construct the sequential input of the RNN model. Another interesting future direction may involve the implementation of RNN-based HSIC frameworks in a real multitemporal HSI context.

X. Strategies for Limited Labeled Samples

Although DNNs have been successfully exploited for the task of HSIC, however, they require a considerably large amount of labeled training data. However, as discussed earlier, the collection of labeled HSI is very critical and expensive due to numerous factors that either demand human experts or exploration of real-time scenarios. The limited availability of labeled training data hinders classification performance. To overcome the aforesaid issue, many effective strategies have been proposed in the literature. In this section, we will briefly discuss some of these strategies while focusing on active learning (AL) algorithms.

A. Data Augmentation

To combat the issue of limited training samples, DA is proven to be an effective tool for HSIC. It generates new samples from the original training samples without introducing additional labeling costs. DA approaches can be categorized into two main strategies as i) data wrapping and ii) oversampling [257]. Data wrapping usually encodes several invariances (translational, size, viewpoint, and/or illumination) by conducting geometric and color-based transformations while preserving the labels, and oversampling-based augmentation methods inflate the training data by generating synthetic samples based on original data distributions. Oversampling techniques include mixture-based instance generation, feature space augmentations [257], and generative adversarial networks (GANs) [258].

Referring to HSIC literature, several DA-based frameworks have been employed to improve the classification performance by avoiding potential overfitting, which is generally caused by the limited availability of training data. For instance, Yu *et al.* [259] enhanced the training data by using three DA operations (flip, rotate, and translation), and then this enhanced data is exploited to train CNN for HSIC. The work [260] presented a comprehensive comparison of various extensively utilized HSI DA techniques and proposed a pixel-block pair-based DA that utilized both spectral and spatial information of HSI to synthesize new instances, to train a CNN model for HSIC. The work [261] compared the classification performance of a combination of CNN and AL with and without DA techniques and demonstrated that the DA leads to higher classification accuracies. Similarly, in another comparison [262], DA-based CNN exhibited a 10% increase in HSIC accuracy when compared to a PCA-based CNN model.

The above-discussed methods utilize offline DA techniques that increase the training data by creating new instances during/before the training process of a model. Recently, a novel DA framework for HSI is proposed in [263], which, rather than inflating the training data, generates the samples at test time, and a DNN trained over original training data along with a voting scheme is used for the final class label. To improve the generalization capability of DNN models, the work [263] also proposed two fast DA techniques for high-quality data syncretization. A similar PCA-based online DA strategy is proposed in [264], which also synthesizes new instances during the inference, instead of training.

B. Semi-Supervised/Unsupervised Learning

SSL approaches learn data distribution by jointly exploiting both labeled and unlabeled data. These techniques expand the training data by utilizing unlabeled samples along with labeled ones in order to construct a relationship between feature space and class labels. Several SSL-based HSIC frameworks have been proposed in the literature that can mainly be categorized as follows: i) Cotraining, ii) self-training, iii) GANs, iv) graph-based SSL models, and v) semi-supervised SVM. A recent comprehensive survey on these SSL techniques can be found in [265]. Moreover, another in-depth survey of SSL approaches is also presented in [266].

The SSL-based HSIC techniques are briefly summarized in [267], where authors also made a detailed comparison of these methods. The method presented in [251] used pseudo or cluster-labeled samples to pretrain a CRNN for HSIC and small-sized labeled data is used to fine-tune the network. Similarly, Kang *et al.* [155] proposed a semi-supervised HSIC framework that exploits PCA and extended morphological attribute profiles to extract pseudo-labeled samples which are fed to a CNN-based deep feature fusion network.

The work [268] proposed a dual strategy cotraining approach based on spectral and spatial features of HSI. Similarly, Zhou *et al.* [269] separately pretrained two SAEs, one using spectral and the other using spatial features of HSI, and fine-tuning is achieved via a cotraining approach. Li *et al.* [270] proposed a region information-based self-training approach to enhance the training data. A graph-based self-training framework was developed in [271], where initial sampling is achieved through subtractive clustering. Recently, Wu *et al.* [156] improved the HSIC performance by pseudo-labeling the unlabeled samples through a clustering-based self-training mechanism and regulating the self-training by employing spatial constraints.

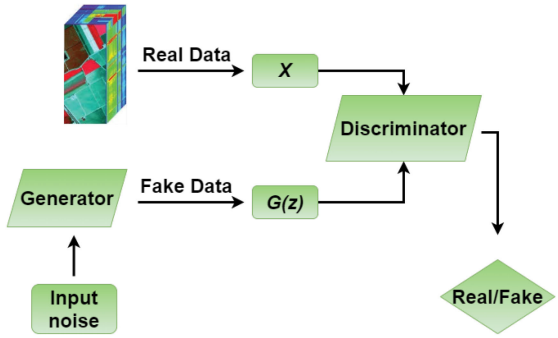


Fig. 13. A general architecture of generative adversarial network (GAN).

C. Generative Adversarial Network

GAN proposed by Goodfellow *et al.* [272] is composed of two neural networks, one is known as a generator and the other is known as discriminator (Fig. 13). GANs can learn to replicate the samples by exploiting the data distribution details. The work [273] proposed a spectral feature-based GAN for SSL-based HSIC.

Similarly, He *et al.* [274] proposed a GAN-based spectral–spatial HSIC framework. Zhu *et al.* [275] developed CNN-based 1-D-GAN and 3-D-GAN architectures to enhance the classification performance. A 1-D customized GAN is used to generate the spectral features [276], which is further used by CNN for FE, and then majority voting is performed HSIC. Very recently, Feng *et al.* [277] introduced a spatial–spectral multiclass GAN (MSGAN) which utilizes two generators to produce spatial and spectral information with the help of multiple adversarial objectives. To address the data imbalance problem for HSIC, Zhong *et al.* [278] proposed a new semi-supervised model which combines GAN with conditional random fields.

Similarly, Wang *et al.* [279] investigated a Caps-TripleGAN model, which effectively generates new samples using a 1-D structure Triple GAN (TripleGAN) and classifying the generated HSI samples using the capsule network (CapsNet). The work [280] proposed to utilize a 3-D-CNN-based generator network and a 3-D deep residual network-based discriminator network for HSIC. To learn high-level contextual features, combination of both capsule network and convolutional LSTM (ConvLSTM)-based discriminator model has been proposed in [281] for HSIC.

The work [282] proposed to address the scarcity of training examples by utilizing a GAN model where the performance of the discriminator is further improved by an auxiliary classifier to produce more structurally coherent virtual training samples. Besides this, to enhance the model performance, Roy *et al.* [283]

proposed a generative adversarial minority oversampling-based technique for addressing the long-standing problem of class-wise data imbalanced imposed by HSIC.

D. Transfer Learning

Transfer learning (TL) enhances the performance of a model by using prior knowledge of a relevant primary task to perform a secondary task. In other words, information extracted from the relevant source domain is transferred to the target domain to learn unseen/unlabeled data. Therefore, TL can be effectively employed in domains with insufficient or no training data. Based on the availability of labeled training instances, TL frameworks can further be categorized as supervised or unsupervised TL. Generally, both source and target domains are assumed to be related but not exactly similar. However, they may follow different distributions as in the case of HSIC where categories of interest are the same but data in two domains may vary due to different acquisition circumstances.

In DNN-based HSIC, the model learns features in a hierarchical manner, where lower layers usually extract generic features, when trained on various images. Therefore, the features learned by these layers can be transferred to learn a new classifier for the target dataset. For instance, Yang *et al.* [284] pretrained to a two-branch spectral–spatial CNN model with an ample amount of training data from other HSIs and then applied the lower layers of the pretrained model to the target network for the robust classification of target HSI. To learn the target-specific features, higher layers of the target network are randomly initialized and the whole network is fine-tuned by utilizing limited labeled instances of target HSI. Similarly, Windrim *et al.* [285] proposed a suitable method to pretrain and fine-tune a CNN network to utilize it for the classification of new HSIs. The study [286] combined DA and TL approaches to combat the shortage of training data in order to improve HSIC performance.

As discussed before, data in source and target domain may vary in many aspects; for instance, in the case of HSIs, the dimensions of two HSIs may vary due to the acquisition from different sensors. Handling such cross-domain variations and transferring the knowledge between them is known as heterogeneous TL (a detailed survey of such methods can be found in [287]). In HSIC literature, several works have been proposed to bridge the gap for transferring the knowledge between two HSIs, with varying dimensions and/or distributions.

For example, Lin *et al.* [288] proposed an effective heterogeneous TL-based HSIC framework that works well with both homogeneous and heterogeneous HSIs, and Li *et al.* [289] used an iterative reweighting mechanism-based heterogeneous TL for HSIC. Similarly, a recent work [290] proposed a band selection-based TL approach to pretrain a CNN, which retains the same number of dimensions for various HSIs. Furthermore, Lin *et al.* [291] proposed an unsupervised TL technique to classify completely unknown target HSI and de Lima and Marfurt [292] demonstrate that the networks trained on natural images can enhance the performance of TL for remote sensing data classification as compared to the networks trained from scratch using smaller HSI data.

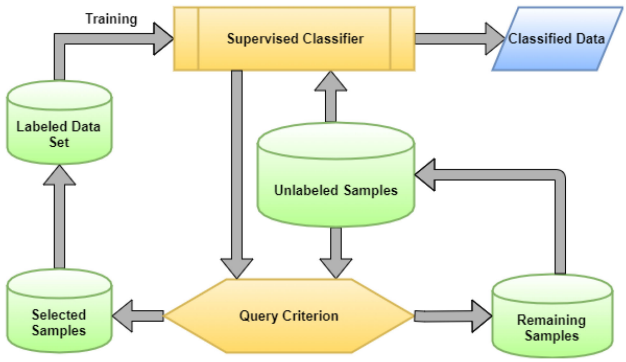


Fig. 14. A general overview of active learning.

E. Active Learning

AL iteratively enhances the predictive performance of a classifier by actively increasing the size of training data, for each training iteration, by utilizing an unlabeled pool of samples. In each iteration, AL enhances the training dataset by actively selecting the most valuable instances from the pool of unlabeled data and an oracle (human- or machine-based) assigns the true class labels to these instances. Finally, these useful instances are added to the existing training dataset and the classifier is retrained on this new training dataset. The process continues until a stopping criterion, that may be the size of the training dataset, the number of iterations, or the desired accuracy score, is achieved. A general framework of AL is illustrated in Fig. 14.

The selection of the most useful/effective samples is made in such a way that the samples should be informative and representative of the overall input distribution in order to improve accuracy. Based on the criteria of adding new instances to the training set, AL frameworks can be designated as either stream-based or pool-based. In stream-based selection, one instance at a time is drawn from an actual set of unlabeled samples and the model decides whether to label it or not based on its usefulness. While in pool-based strategy, samples are queried from a pool/subset of unlabeled data based on ranking scores computed from various measures to evaluate the sample's usefulness.

The work [293] found that streamed-based selection gives poorer learning rates as compared to pool-based selection as the former tends to query extra instances. In pool-based selection, it is important to incorporate diversity in the pool of samples, in order to avoid redundancy within the pool of samples. Generally, the following three aspects are focused on while selecting/querying the most valuable samples: heterogeneity behavior, model's performance, and representativeness of samples. A brief introduction of these sampling approaches is given below.

1) Heterogeneity-Based Selection: These approaches select the samples that are more heterogeneous to the already seen instances with respect to model diversity, classification uncertainty, and contention between a committee of various classifiers. Uncertainty sampling, expected model change, and query-by-committee are examples of heterogeneity-based models.

- *Uncertainty Sampling:* In this approach, the classifier iteratively tries to query the label of those samples for which it

is most uncertain while predicting the label. The selection of new instances is based on ranking scores against a specified threshold and the instances with scores closest to that threshold are queried for labels. One simple example of such a scheme could be implementing the probabilistic classifier on a sample in a scenario of binary classification and querying its label if the predicted class probability is close to 0.5.

- *Query-by-Committee:* Such heterogeneity-based approaches perform the sampling process based on the dissimilarities in the predictions of various classifiers trained on the same set of labeled samples. A committee of various classifiers trained on the same set of training data is used to predict the class labels of unlabeled samples, and the samples for which classifiers differ more are selected for querying labels. The committee of different classifiers can either be built by using ensemble learning algorithms like Bagging and Boosting [294] or by changing the model parameters [295]. Generally, a less number of diverse classifiers is adequate for constructing a committee [294], [296].
- *Expected Model Change:* Such a heterogeneity-based approach chooses the instances which result in a significant change from the current model in terms of the gradient of the objective function. Such techniques attempt to query the label for those instances that are considerably different from the current model. These sampling techniques only fit the models which follow gradient-based training procedures/optimization.

2) Performance-Based Selection: Such methods consider the effect of adding queried samples to the model performance. They try to optimize the performance of the model by reducing variance and error. There are two types of performance-based sampling.

- *Expected Error Reduction:* This approach is interrelated to uncertainty sampling in such a way that uncertainty measures maximize the label uncertainty of the sample to be queried for the label while expected error reduction reduces the label uncertainty of the queried sample. Referring to the already discussed example of the binary classification problem, the expected error reduction approach would choose the samples with a probability far away from 0.5 in order to reduce the error rate. Such techniques are also known as the greatest certainty models [295].
- *Expected Variance Reduction:* Reducing the variance of the model is guaranteed to reduce future generalization error [297]. Therefore, expected variance reduction techniques attempt to indirectly reduce the generalization error by minimizing the model variance. Such approaches query the instances that result in the lowest model variance. The Fisher information ratio is a well-known variance minimization framework.

3) Representativeness-Based Selection: Heterogeneity-based models are prone to include outlier and controversial samples but performance-based approaches implicitly avoid such samples by estimating future errors. Representative sampling tends to query such instances that are representative of the overall input distribution; hence, avoid outliers and unrepresentative samples. These approaches weigh the dense input region to a higher degree while the querying process. Density-weighted techniques like information density are examples of representativeness sampling approaches that consider the representativeness of samples along with heterogeneity behavior, and are also known as hybrid models [295].

Recently, AL has been intensively utilized in HSIC. Liu *et al.* [298] proposed a feature-driven AL framework to define a well-constructed feature space for HSIC. Zhang *et al.* [299] proposed a RF-based semi-supervised AL method that exploits spectral–spatial features to define a query function to select the most informative samples as target candidates for the training set.

Spatial information has been intensively exploited in many AL-based HSIC. For instance, Guo *et al.* [300] presented an AL framework that splices together the spectral and spatial features of superpixels. Similarly, Xue *et al.* [301] considered the neighborhood and superpixel information to enhance the uncertainty of queried samples. In recent work, Bhardwaj *et al.* [302] exploited the attribute profiles to incorporate spatial information in an AL-based HSIC framework.

Batch-mode AL frameworks have been widely employed to accelerate the learning process. Such approaches select a batch of samples, in each iteration, to be queried for a label. Therefore, the diversity of the samples is extremely critical in batch mode AL techniques in order to avoid redundancy. A multicriteria batch-mode AL method proposed by Patra *et al.* [303] defines a novel query function based on diversity, uncertainty, and cluster assumption measures. These criteria are defined by exploiting the properties of KNN, SVM, and K-means clustering, respectively, and finally, genetic algorithms are used to choose the batch of most effective samples. Similarly, Zhang and Crawford [304] proposed a regularized multimetric batch-mode AL framework for HSIC that exploits various features of HSI.

A multiview AL (MVAL) framework was proposed in [305] that analyzes the object from various views and measures the informativeness of the sample through multiview intensity-based query criteria. Similarly, Pradhan *et al.* [306] also exploited the concept of multiview learning using the Fisher discriminant ratio to generate multiple views. In another work, Zhang *et al.* [307] proposed a novel adaptive MVAL framework for HSIC, which jointly exploits the spatial and spectral features in each view. Recently, Li *et al.* [308] proposed an MVAL technique that utilizes pixel-level, subpixel-level, and superpixel-level details to generate multiple views for HSIC. Moreover, the proposed method exploits joint posterior probability estimation and dissimilarities among multiple views to query the representative samples.

In the HSIC literature, several works have combined the AL and DNN. For instance, Sun *et al.* [309] joined AE with AL technique and Liu *et al.* [310] proposed a DBN-based AL framework for HSIC. Similarly, Haut *et al.* [311] coupled Bayesian CNN with AL paradigm for spectral–spatial HSIC. Recently,

Rochac *et al.* [261] proposed a CNN-based AL framework to better exploit the unlabeled samples for HSIC.

Many works integrated AL with TL for HSIC. For example, Lin *et al.* [312] proposed an AL-based TL framework that extracts the salient samples and exploits high-level features to correlate the source and target domain data. Another work [313] proposed a stacked sparse AE-based active TL technique that jointly utilizes both spectral and spatial features for HSIC. Another work [314] combined domain adaptation and AL methods based on multiple kernels for HSIC.

AL-based HSIC offers some sophisticated frameworks to enhance the generalization capabilities of models. For instance, Ahmad *et al.* [35] proposed a fuzziness-based AL method to improve the generalization performance of discriminative and generative classifiers. The method computes the fuzziness-based distance of each instance and estimated class boundary, and the instances having greater fuzziness values and smaller distances from class boundaries are selected to be the candidates for the training set. Recently, Ahmad *et al.* [315] proposed a nonrandomized spectral–spatial AL framework for multiclass HSIC that combines the spatial prior fuzziness approach with MLR via a splitting and augmented Lagrangian classifier. The authors also made a comprehensive comparison of the proposed framework with state-of-the-art sample selection methods along with diverse classifiers.

XI. Experimental Evaluation

The most research-oriented works published in the literature present a comprehensive experimental evaluation to highlight the pros and cons of the work/s proposed. However, to some extent, these works may have chosen different experimental settings; for instance, training, validation, and test samples may have the same number or percentage but the samples may be different as these samples are normally chosen randomly. Therefore, to make a fair comparison among different works proposed in the literature, one must need to have the same experimental settings.

These experimental settings include the same samples (geographical locations should remain the same for all chosen models no the different ones) and the number of samples should have been selected for each round of training in the cross-validation process. Normally, these samples have been chosen randomly, and thus high likely, they may be different for different models if the models are executed at different times.

The other issue with most of the literature proposed in recent years is overlapping between training/test samples, i.e., training/validation samples have been randomly selected (including or excluding the above point) for training and validation; however, the entire dataset has been passed at a testing phase, which leads to a highly biased model (as the training samples have already been seen by the model) and produces high accuracy. Thus, in this work, the training/test samples are though chosen randomly (because all the models have been executed at the same time); however, the above point has been taken seriously and the intersection among these samples remains empty.

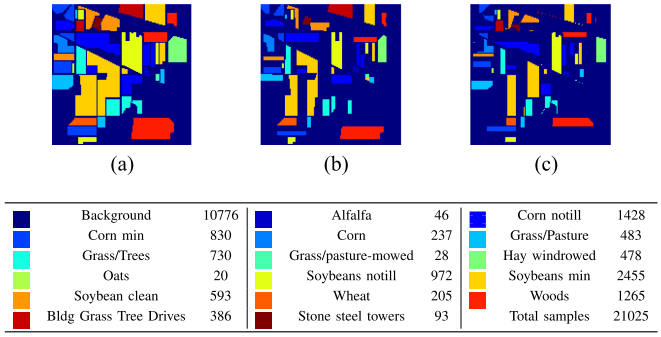


Fig. 15. Type associated with the land-cover classes and number of available samples in the Indian Pines (IP) dataset. Moreover, spatially disjoint training and test samples for the IP dataset are also presented.

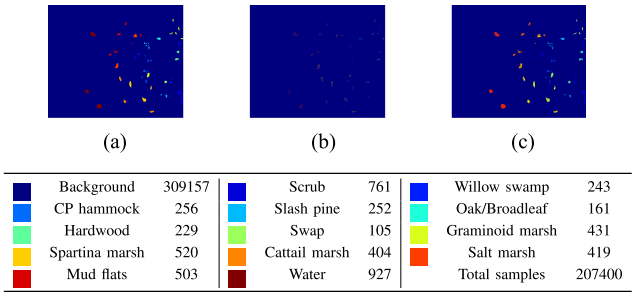


Fig. 16. Type associated with the land-cover classes and number of available samples in the Kennedy Space Center (KSC) dataset. Moreover, spatially disjoint training and test samples for the KSC dataset are also presented.

A. Experimental Datasets

The Indian Pines (IP) dataset was gathered by the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) [316] over the Indian Pines test site in North-Western Indiana. It contains 224 spectral bands within a wavelength range of 400–2500 nm. The 24 null and corrupted bands have been removed. The spatial size of the image is 145×145 pixels, and it comprises 16 mutually exclusive vegetation classes. The spatial resolution is 20 m per pixel (MPP). The detailed class description and ground-truth maps are presented in Fig. 15(a). Moreover, the disjoint training/test sample maps are presented in Fig. 15(b) and (c).

The Kennedy Space Center (KSC) dataset was gathered in 1996 by AVIRIS [316], with wavelengths ranging from 400–2500 nm. The image has 512×614 pixels and 176 spectral bands after removal of some low signal-to-noise ratio bands. The KSC dataset comprises 5202 labeled samples, with a total of 13 upland and wetland classes. The detailed class description and ground truth maps are presented in Fig. 16(a). Moreover, the disjoint training/test sample maps are presented in Fig. 16(b) and (c).

The University of Pavia (UP) dataset was acquired by the Reflective Optics System Imaging Spectrometer (ROSIS) sensor during a flight campaign over the university campus at Pavia, Northern Italy [317]. It consists of 610×340 pixels with 103 spectral bands in the wavelength range from 430 to 860 nm and 2.5 MPP. It comprises nine urban land-cover classes. The detailed class description and ground-truth maps are presented

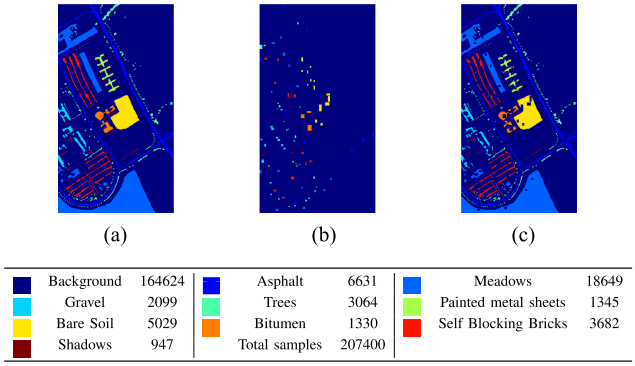


Fig. 17. Type associated with the land-cover classes and number of available samples in the Pavia University (PU) dataset. Moreover, spatially disjoint training and test samples for the PU dataset are also presented.

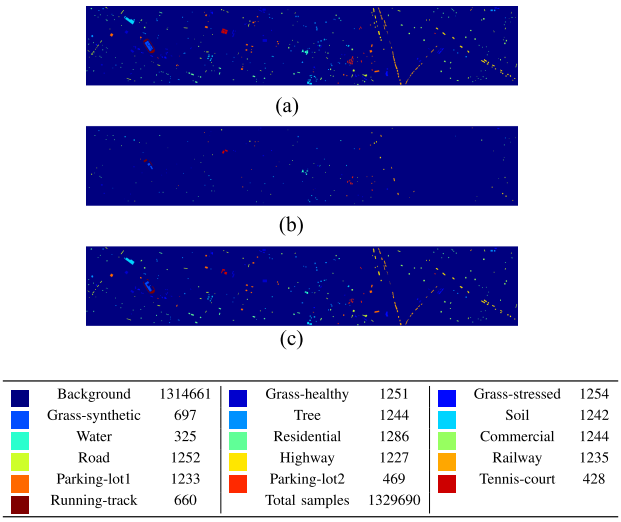


Fig. 18. Type associated with the land-cover classes and number of available samples in the Houston (UH) dataset.

in Fig. 17(a). Moreover, the disjoint training/test sample maps are presented in Fig. 17(b) and (c).

The IEEE Geoscience and Remote Sensing Society published the University of Houston (UH) dataset—collected by the Compact Airborne Spectrographic Imager (CASI)—in 2013 [318], as part of its data fusion contest. It is composed of 340×1905 pixels with 144 spectral bands. The spatial resolution of this dataset is 2.5 MPP with a wavelength ranging from 0.38 to 1.05 μm. Finally, the ground-truth comprises 15 different land-cover classes. The detailed class description and ground-truth maps are presented in Fig. 18(a) and disjoint training/test samples maps are presented in Fig. 18(b) and (c).

The University of Trento (UT) dataset was gathered by the using AISA eagle sensor over the rural regions in the south of Trento, Italy. The HSI contains 63 spectral bands within a wavelength of range 0.42 to 0.99 μm [319]. The scene has 600×166 pixels, which comprises of six mutually exclusive vegetation land-cover classes where the spectral resolution is 9.2 nm, and the spatial resolution is 1 MPP. In addition, the available samples are divided into disjoint training and test samples of six classes, and Fig. 19 lists the information about the per class number of samples for six different land-covers.

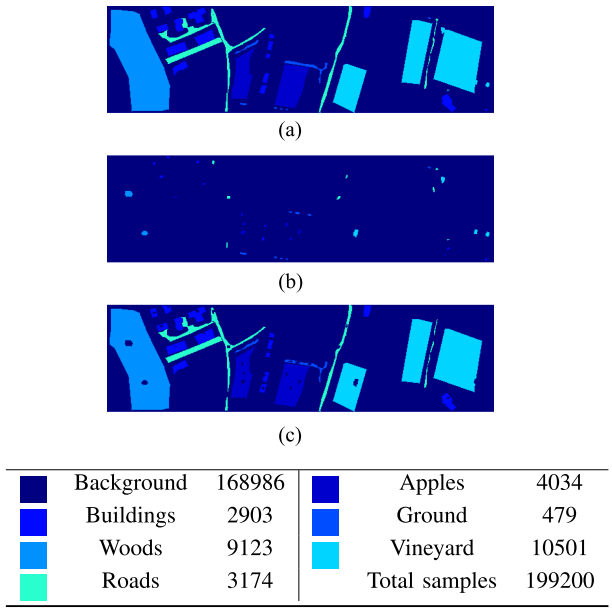


Fig. 19. Type associated with the land-cover classes and number of available samples in the University of Trento (UT) dataset.

TABLE I
SUMMARY OF THE HSI DATASETS USED FOR EXPERIMENTAL EVALUATION

—	IP	PU	KSC	UH	UT
Year	1992	2001	1996	2013	
Source	AVIRIS	ROSIS-03	AVIRIS	CASI	AISA
Spatial	145×145	610×610	512×614	340×1905	600×166
Spectral	220	115	176	144	63
Wavelength	$400 - 2500$	$430 - 860$	$400 - 2500$	$0.35 - 1.05$	$0.42 - 0.99$
Samples	21025	207400	314368	1329690	199200
Classes	16	9	13	15	6
Sensor	Aerial	Aerial	Aerial	Aerial	Aerial
Resolution	20 m	1.3 m	10 nm	2.5 mpp	1 mpp

Table I provides a summary description of each dataset used in the following experiments, whereas Table II enlists the numbers of disjoint samples (i.e., train/test samples selected from each class) used for all the experimental results. Note that the number of train/test (i.e., percentage) samples and geographical locations of train/test samples remain the same for all experimental methods (competing methods).

B. Experimental Results on Disjoint Train/Test Samples

To strengthen the ideas highlighted in this survey and to make the claims valid, the main contributions made in recent years include MLR, SVM, MLP, RNN, LSTM, GRU, CNN-1D, CNN-2D, CNN-3D, and MorphCNN, which have been considered to compare the experimental results. Some of the representative works for each of the above methods are as follows: cloud implementation of logistic regression for HSIC [320], [326], [327] (MLR), classification of hyperspectral remote sensing images with SVM [321], deep RNNs for HSIC [70], LSTM [322]; on the properties of neural machine translation: encoder–decoder approaches [323] (GRU), deep CNNs for HSIC (CNN1D) [217], deep supervised learning for hyperspectral data classification through CNNs (CNN2D) [324], 3-D DL approach for remote sensing image classification (CNN3D)[325], morphological CNNs for HSIC (MorphCNN) [195], and MLP [69].

TABLE II
NUMBER OF DISJOINT TRAIN/TEST SAMPLES USED FOR THE EXPERIMENTAL RESULTS

IP Data				KSC Data				PU Data				UH Data				UT Data			
Class	Land Cover	TRS	TES	Class	Land Cover	TRS	TES	Class	Land Cover	TRS	TES	Class	Land Cover	TRS	TES	Class	Land Cover	TRS	TES
1		29	25	1		114	602	1		548	6083	1		198	1053		Apples	129	3905
2		762	675	2		36	207	2		540	18109	2		190	1064		Buildings	125	2778
3		435	404	3		38	218	3		392	1707	3		192	505		Ground	105	374
4		146	99	4		38	214	4		524	2540	4		188	1056		Woods	154	8969
5		232	274	5		24	137	5		265	1080	5		186	1056		Vineyard	184	10317
6		394	354	6		34	195	6		532	4497	6		182	143		Roads	122	3052
7		16	2	7		16	89	7		375	955	7		196	1072				
8		235	250	8		65	366	8		514	3168	8		191	1053				
9		10	10	9		78	442	9		231	716	9		193	1059				
10		470	503	10		61	343					10		191	1036				
11		1424	1065	11		63	356					11		181	1054				
12		328	282	12		75	428					12		192	1041				
13		132	80	13		139	788					13		184	285				
14		728	545									14		181	247				
15		291	99									15		187	473				
16		57	44																

TRS and TES stands for disjoint train and test samples, respectively.

TABLE III
CLASSIFICATION RESULTS OBTAINED BY RF [75], MLR [320], SVM [321], MLP [69], RNN [70], LSTM [322], GRU [323], CNN-1D [217], CNN-2D [324], CNN-3D [325], HYBRIDSN [212], AND MORPHCNN [195] ON THE DISJOINT TRAIN–TEST DATASET FOR THE PU SCENE

Class	RF [75]	MLR [321]	SVM [322]	MLP [69]	RNN [70]	LSTM [323]	GRU [324]	CNN-1D [218]	CNN-2D [325]	CNN-3D [326]	HybridSN [213]	MorphCNN [196]
1	89.98±0.15	77.68±0.0	82.23±0.0	84.53±1.89	83.08±3.3	82.63±2.39	77.25±6.92	87.18±2.11	93.4±1.89	85.66±4.0	89.74±5.19	**94.52**±1.9
2	74.39±0.01	58.79±0.01	65.81±0.0	75.13±2.4	67.9±2.92	78.74±1.99	80.1±5.12	89.64±2.53	96.84±1.93	95.88±1.71	81.78±3.15	**97.12**±8.71
3	38.42±0.13	67.21±0.02	66.72±0.0	68.37±5.17	65.17±7.99	60.73±11.0	54.79±14.82	71.1±5.98	65.48±13.94	68.11±6.47	82.88±1.54	**85.08**±4.53
4	98.24±0.05	74.27±0.05	**97.77**±0.0	93.5±2.32	90.72±2.56	97.1±1.22	92.05±2.31	95.32±1.49	95.55±2.14	97.02±0.83	83.66±4.58	97.0±1.03
5	95.98±0.04	98.88±0.04	99.37±0.0	99.37±0.08	99.23±0.09	99.28±0.08	99.51±0.12	99.48±0.26	98.03±0.92	98.9±0.56	**99.94**±0.04	99.25±0.22
6	51.43±0.19	93.53±0.02	91.62±0.0	89.94±4.14	85.07±3.14	65.94±5.92	74.86±11.38	88.28±2.33	80.52±9.39	68.85±11.29	72.43±13.23	**93.92**±3.88
7	80.63±0.36	85.08±0.05	87.36±0.0	87.2±3.05	82.94±3.79	84.95±4.02	90.17±3.9	86.77±3.38	89.29±9.48	73.09±9.53	**96.16**±1.88	84.98±10.74
8	97.64±0.14	87.58±0.01	90.46±0.0	90.37±1.24	85.85±4.97	88.89±7.83	90.42±4.39	90.43±3.34	94.5±5.44	95.21±1.69	92.80±0.90	**96.62**±2.21
9	94.92±0.05	**99.22**±0.05	93.71±0.0	98.44±1.17	94.52±4.79	98.29±1.47	93.51±7.93	97.33±3.31	95.8±0.76	93.54±1.76	94.04±3.99	97.05±0.46
OA	77.44±0.06	72.23±0.0	77.8±0.0	82.05±0.88	77.07±0.95	80.38±0.52	80.7±0.56	89.09±0.97	92.55±1.02	89.43±1.37	84.18±1.40	**95.51**±0.66
AA	80.18±0.06	82.47±0.01	86.12±0.0	87.43±1.03	83.83±0.72	84.06±0.74	83.63±2.03	89.5±1.03	89.94±1.37	86.25±1.98	88.16±1.94	**93.95±0.96**
k(x100)	70.44±0.07	65.44±0.0	72.06±0.0	76.89±1.07	70.84±1.04	74.32±0.68	74.76±1.02	85.5±1.22	89.9±1.42	85.61±1.94	79.13±1.42	**93.95**±0.88

TABLE IV
CLASSIFICATION RESULTS OBTAINED BY RF [75], MLR [320], SVM [321], MLP [69], RNN [70], LSTM [322], GRU [323], CNN-1D [217], CNN-2D [324], CNN-3D [325], HYBRIDSN [212], AND MORPHCNN [195] ON THE DISJOINT TRAIN–TEST DATASET FOR THE IP SCENE

Class	RF [75]	MLR [321]	SVM [322]	MLP [69]	RNN [70]	LSTM [323]	GRU [324]	CNN-1D [218]	CNN-2D [325]	CNN-3D [326]	HybridSN [213]	MorphCNN [196]
1	85.33±1.88	80.0±0.0	88.0±0.0	73.6±7.42	58.4±4.8	89.6±1.96	77.6±9.33	80.8±12.75	73.64±14.77	48.18±22.84	82.66±13.59	**92.27**±3.55
2	55.11±0.32	81.48±0.0	80.0±0.0	81.45±1.07	75.5±1.48	82.22±1.26	81.1±2.77	79.38±4.16	83.12±6.12	**85.12**±7.88	82.17±2.64	84.05±8.03
3	22.77±0.20	54.11±0.12	69.55±0.0	64.55±2.85	63.37±1.93	64.16±5.44	70.35±1.36	74.26±6.12	**81.98**±3.89	77.22±13.04	76.73±4.02	79.34±3.45
4	13.13±1.64	38.38±0.0	48.48±0.0	47.07±10.41	29.49±5.4	**55.35**±10.62	53.33±11.15	31.92±11.55	45.39±6.36	50.11±10.04	33.33±3.59	52.14±6.24
5	41.60±0.78	**91.97**±0.0	87.23±0.0	86.94±1.07	87.59±1.5	89.27±0.68	88.4±0.85	90.73±1.07	89.11±5.55	80.28±6.52	81.14±8.89	91.66±1.69
6	94.06±0.23	94.63±0.0	96.33±0.0	95.93±0.97	95.31±0.83	96.39±0.87	96.38±1.14	**96.39**±0.9	95.02±5.68	89.81±4.03	97.36±2.54	95.74±2.28
7	0.0±0.0	0.0±0.0	**50.0**±0.0	10.0±20.0	0.0±0.0	0.0±0.0	0.0±0.0	0.0±0.0	0.0±0.0	0.0±0.0	0.0±0.0	0.0±0.0
8	91.33±0.18	**100.0**±0.0	**100.0**±0.0	99.84±0.2	99.52±0.3	99.2±0.91	99.12±0.64	99.84±0.32	99.96±0.13	95.96±6.78	96.53±3.78	**100.0**±0.0
9	40.0±0.0	0.0±0.0	50.0±0.0	**80.0**±15.49	56.0±10.2	76.0±8.0	66.0±4.9	50.0±8.94	26.66±15.87	77.78±21.66	66.66±24.94	44.44±19.88
10	26.83±1.26	66.76±0.08	76.54±0.0	75.35±5.02	71.13±5.93	81.51±2.76	78.53±4.64	**81.83**±4.69	77.44±8.99	77.9±6.2	74.35±9.46	80.77±3.77
11	81.06±0.49	84.13±0.0	87.7±0.0	83.19±1.51	78.86±1.45	80.4±2.43	82.29±1.82	80.39±3.65	**89.4**±5.47	82.73±3.81	79.18±4.92	88.54±5.03
12	28.95±0.44	66.31±0.0	77.3±0.0	78.58±2.95	71.91±5.07	76.31±1.39	83.19±1.16	84.75±7.5	87.72±3.06	82.64±14.49	71.04±4.11	**88.46**±4.26
13	86.25±1.02	95.0±0.0	97.5±0.0	**98.0**±0.61	97.0±1.7	97.25±0.94	97.75±0.94	97.75±0.5	95.28±4.74	89.72±6.89	96.25±4.44	87.64±3.43
14	91.07±0.87	90.64±0.0	91.38±0.0	92.92±1.48	90.28±1.09	94.13±1.18	92.88±1.79	93.32±2.34	**98.94**±0.55	98.31±1.41	91.68±4.33	98.82±1.01
15	10.10±0.0	89.9±0.0	80.81±0.0	87.88±3.78	75.56±6.43	90.71±2.34	**93.54**±1.64	89.9±4.78	82.02±14.83	55.17±27.57	45.45±21.39	69.44±15.86
16	71.96±3.86	**97.73**±0.0	97.73±0.0	87.27±4.45	88.64±4.31	94.09±2.32	95.45±2.49	96.82±2.32	82.0±6.69	82.5±12.5	84.09±4.90	84.0±4.21
OA	60.80±0.14	80.33±0.02	84.12±0.0	82.95±0.23	79.07±0.33	83.55±0.39	84.2±0.21	84.0±0.28	87.25±1.03	83.6±1.41	80.86±1.74	**87.45**±1.01
AA	52.47±0.21	70.69±0.01	**79.91**±0.0	77.66±1.98	71.16±0.87	79.16±0.75	78.49±0.36	76.76±0.75	75.48±2.12	73.34±3.46	72.41±2.32	77.33±1.56
k(x100)	54.41±0.18	77.47±0.02	81.87±0.0	80.56±0.26	76.12±0.4	81.27±0.44	82.01±0.26	81.81±0.35	85.48±1.15	81.36±1.62	78.24±1.98	**85.75**±1.14

To some extent, all the aforesaid works are based on convolutional and recurrent networks and are evaluated on four benchmark HSI datasets, namely IP, PU, KSC, Houston Scene, and the University of Toronto. This survey only pays attention to the robustness of all these models while considering the small sample size of training data to classify HSI for joint spatial–spectral classification.

Here we have enlisted the experimental results with detailed discussion on the obtained results. The obtained accuracies for disjoint training and test samples are shown in Tables III–VII and Figs. 20–24. All the results shown in the Tables and Figures are obtained using the 10-cross-validation process to compute the overall, average, and kappa (κ) accuracy for comparison purposes. For instance, let us assume the case of Pavia University results, for this particular case; the work [195] has the highest average, overall, and kappa (κ) accuracies which are 95.51%, 93.95%, and 93.95%, respectively, in comparison with the average, overall, and kappa (κ) accuracies for other comparative works; 92.55%, 89.94%, 89.9% for [324], 89.43%, 86.25%, 85.61% for [325], 89.09%, 89.5%, 85.5% for [217], 82.05%, 87.43%, 76.89% for [69], 80.38%, 83.63%, 74.76% for [323], 80.38%, 84.06%, 74.32% for [322], 77.8%, 86.12%, 72.06% for [321], 77.07%, 83.83%, 70.84% for [70], and 72.23%, 82.12%, 65.44% for [320]. Similar observations can be made of the other experimental datasets. The higher accuracies are emphasis in bold in Tables III–VIII.

The comparative methods mostly misclassify the samples having similar spatial structures (i.e., Meadows and Bare Soil classes for Pavia University dataset) as shown in Table III and Fig. 20. Moreover, the overall accuracy for Grapes Untrained is lower than the other classes due to the reasons mentioned above. In a nutshell, one can say that higher accuracy can be achieved

TABLE V
CLASSIFICATION RESULTS OBTAINED BY RF [75], MLR [320], SVM [321], MLP [69], RNN [70], LSTM [322], GRU [323], CNN-1D [217], CNN-2D [324], CNN-3D [325], HYBRIDSN [212], AND MORPHCNN [195] ON THE DISJOINT TRAIN–TEST DATASET FOR THE UH SCENE

Class	RF [75]	MLR [321]	SVM [322]	MLP [69]	RNN [70]	LSTM [323]	GRU [324]	CNN-1D [218]	CNN-2D [325]	CNN-3D [326]	HybridSN [213]	MorphCNN [196]
1	82.87±0.04	82.24±0.06	82.34±0.0	81.23±0.28	82.22±0.28	**82.76**±0.36	82.58±0.35	82.28±0.98	82.25±0.65	82.1±0.39	82.74±0.29	82.43±0.33
2	82.51±0.38	82.5±0.07	83.36±0.0	82.29±0.55	82.87±0.33	80.19±1.36	81.64±0.71	**91.78**±6.46	84.15±0.28	84.14±0.45	90.91±6.22	84.42±0.19
3	64.09±0.49	99.8±0.0	99.8±0.0	99.72±0.1	99.72±0.2	99.68±0.16	99.88±0.1	**99.92**±0.16	90.31±4.41	77.85±4.8	98.81±0.74	97.21±1.23
4	92.04±0.0	98.3±0.0	**98.96**±0.0	87.58±1.28	93.5±2.02	91.23±1.29	93.22±2.84	94.36±3.12	87.24±3.21	89.24±1.1	83.96±1.24	92.37±0.33
5	99.81±0.07	97.44±0.0	98.77±0.0	97.35±0.49	97.76±0.29	97.65±0.31	97.37±0.16	98.77±0.13	99.51±0.48	98.97±0.59	99.46±0.75	**99.77**±0.53
6	96.27±0.32	94.41±0.0	97.9±0.0	94.55±0.28	95.1±0.0	97.06±1.79	98.32±1.63	95.8±1.88	96.43±2.14	98.91±1.44	98.60±1.97	**99.46**±1.15
7	86.19±0.34	73.37±0.07	77.43±0.0	75.24±2.27	81.4±0.43	78.88±1.0	77.03±2.18	82.78±2.23	86.44±2.18	85.48±1.98	75.62±3.89	**88.07**±1.78
8	41.69±0.23	63.82±0.0	60.3±0.0	57.0±6.97	40.06±1.07	40.11±1.92	53.62±2.97	**75.5**±6.71	70.03±3.96	62.06±3.01	93.16±0.20	73.09±3.5
9	86.02±0.48	70.23±0.04	76.77±0.0	75.58±2.86	76.54±2.96	81.55±4.12	79.06±1.61	81.44±2.0	79.53±6.38	80.81±4.32	81.39±5.24	**84.09**±2.73
10	36.00±0.0	55.6±0.0	61.29±0.0	48.78±2.27	47.44±1.44	47.37±2.29	49.54±2.61	**68.71**±14.55	60.22±4.2	54.75±4.63	76.51±10.40	62.86±3.08
11	64.67±0.16	74.21±0.04	80.55±0.0	76.25±0.46	76.24±0.81	76.38±1.09	80.82±0.71	85.24±2.83	82.93±7.68	66.78±3.34	89.21±5.62	**89.15**±6.86
12	67.27±0.09	70.41±0.0	79.92±0.0	75.31±3.75	76.33±3.09	79.98±3.32	84.15±3.13	89.93±4.29	92.87±3.31	**93.83**±1.92	96.28±2.29	93.02±3.32
13	89.23±0.43	67.72±0.0	70.88±0.0	73.19±2.15	69.12±1.61	71.37±3.54	72.63±3.68	74.88±5.14	86.21±2.65	82.34±2.49	86.78±6.67	**89.61**±1.34
14	100.0±0.0	98.79±0.0	100.0±0.0	99.84±0.32	100.0±0.0	99.11±0.47	99.92±0.16	99.68±0.16	98.92±1.8	96.31±3.67	**100.0**±0.0	99.19±1.3
15	90.06±0.45	95.56±0.0	96.41±0.0	97.8±0.51	97.59±0.47	98.14±0.31	98.22±0.59	98.48±0.24	77.63±2.91	75.85±2.69	**100.0**±0.0	97.04±4.47
OA	75.38±0.06	78.97±0.01	81.86±0.0	78.22±0.36	77.95±0.68	78.16±0.28	80.21±0.27	86.42±1.64	83.27±0.8	80.24±0.55	88.31±1.78	**86.51**±0.71
AA	78.58±0.11	81.63±0.01	84.31±0.0	81.45±0.37	81.06±0.55	81.43±0.32	83.2±0.27	87.97±1.38	84.98±0.74	81.96±0.75	90.23±1.39	**88.78**±0.68
k(x100)	73.49±0.07	77.3±0.01	80.43±0.0	76.55±0.39	76.23±0.71	76.52±0.3	78.66±0.29	85.27±1.77	81.89±0.86	78.62±0.59	87.33±1.92	**85.4**±0.76

TABLE VI
CLASSIFICATION RESULTS OBTAINED BY RF [75], MLR [320], SVM [321], MLP [69], RNN [70], LSTM [322], GRU [323], CNN-1D [217], CNN-2D [324], CNN-3D [325], HYBRIDSN [212], AND MORPHCNN [195] ON THE DISJOINT TRAIN-TEST DATASET FOR THE KSC SCENE

Class	RF [75]	MLR [321]	SVM [322]	MLP [69]	RNN [70]	LSTM [323]	GRU [324]	CNN-1D [218]	CNN-2D [325]	CNN-3D [326]	HybridSN [213]	MorphCNN [196]
1	99.69±0.0	**100.0**±0.00	94.13±0.00	99.18±1.17	87.33±1.73	92.22±0.91	89.44±0.69	99.79±0.14	85.52±15.45	97.17±2.93	**100.0**±0.0	97.63±2.26
2	98.38±0.22	99.03±0.001	0.00±0.00	86.63±7.33	63.12±1.49	81.64±4.17	70.85±0.91	99.19±1.14	67.31±25.41	92.91±2.62	**100.0**±0.0	86.79±9.64
3	99.23±0.57	99.54±0.00	54.59±0.00	84.25±2.70	69.72±2.82	75.38±0.21	78.89±7.79	95.11±3.05	60.09±16.23	81.04±12.05	**99.69**±0.21	98.31±2.37
4	88.16±1.22	99.06±0.002	17.28±0.00	78.97±11.86	47.82±4.20	58.09±1.16	44.08±3.41	77.73±2.81	45.17±9.84	44.54±16.19	**99.53**±0.66	88.94±15.64
5	73.72±0.0	**100.0**±0.00	0.00±0.00	13.38±18.92	68.37±5.63	74.21±5.20	65.21±3.28	80.53±4.85	67.40±12.84	85.15±6.98	98.78±0.34	48.66±34.29
6	88.88±0.24	**100.0**±0.00	0.00±0.00	78.12±8.79	56.24±1.97	65.12±6.03	59.82±2.30	91.97±1.74	65.47±29.63	62.74±15.45	**100.0**±0.0	86.32±16.43
7	**100.0**±0.0	89.88±0.00	0.00±0.00	78.65±3.99	83.52±8.91	90.26±1.40	89.14±3.47	95.13±1.40	77.15±28.34	80.52±17.07	97.75±1.83	97.75±3.17
8	85.51±0.22	**100.0**±0.00	60.10±0.00	89.62±8.25	65.57±2.32	71.40±2.99	69.76±2.53	97.45±0.51	64.75±13.87	71.49±11.96	99.90±0.12	70.76±34.44
9	96.68±0.42	**100.0**±0.00	89.37±0.00	97.59±1.79	88.39±3.35	90.72±2.93	86.72±1.36	99.92±0.11	89.22±8.06	98.94±1.33	**100.0**±0.0	91.93±7.16
10	99.22±0.13	**100.0**±0.00	98.83±0.00	96.50±3.33	92.42±3.12	88.92±3.37	88.53±1.58	99.90±0.13	73.08±23.37	90.67±8.13	**100.0**±0.0	**100.0**±0.00
11	**100.0**±0.0	98.03±0.001	94.94±0.00	98.50±0.86	83.89±2.29	90.26±1.08	84.83±3.64	**100.0**±0.0	87.55±10.06	97.56±1.17	96.34±0.79	**100.0**±0.0
12	97.89±0.19	**99.29**±0.001	89.25±0.00	98.52±0.79	81.31±4.48	87.46±2.05	83.57±4.87	98.36±1.06	82.48±19.17	99.30±0.99	99.06±0.99	97.89±2.06
13	**100.0**±0.0	**100.0**±0.0	**100.0**±0.0	**100.0**±0.0	99.88±0.10	**100.0**±0.0	99.92±0.05	**100.0**±0.0	99.92±0.12	**100.0**±0.0	**100.0**±0.0	**100.0**±0.0
OA	96.17±0.07	99.45±0.001	72.84±0.00	91.76±0.56	81.47±1.17	86.10±0.40	82.76±0.72	97.18±0.18	79.98±13.38	89.71±1.30	**99.48**±0.05	92.76±2.08
AA	94.41±0.08	98.83±0.001	53.73±0.00	84.61±0.62	75.96±1.71	81.97±0.34	77.75±0.70	95.00±0.34	74.24±16.27	84.77±2.60	**99.31**±0.17	89.61±1.60
k(x100)	95.74±0.08	99.40±0.001	69.29±0.00	90.82±0.62	79.33±1.30	84.51±0.45	80.79±0.80	96.86±0.20	77.63±14.99	88.51±1.46	**99.43**±0.06	91.94±2.31

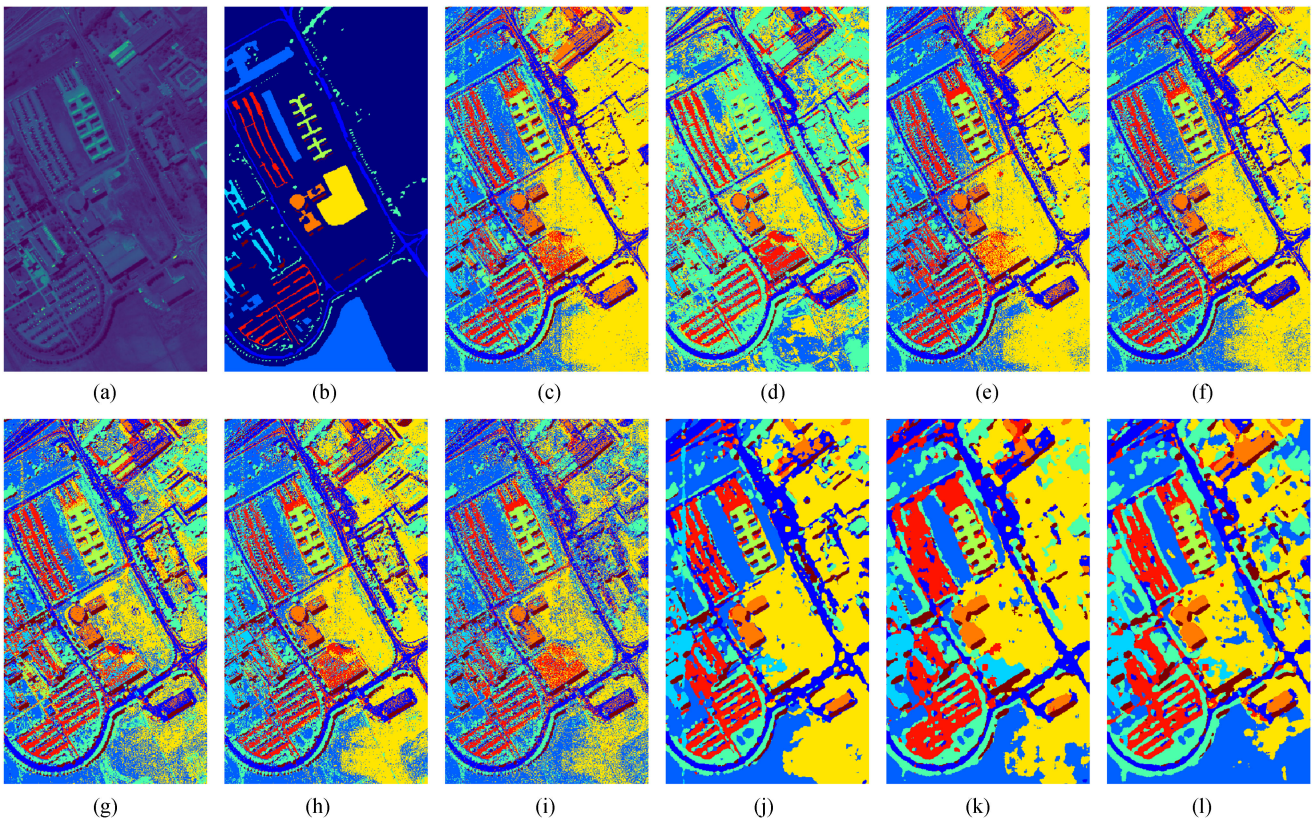


Fig. 20. Classification maps obtained by MLR [320], SVM [321], MLP [69], RNN [70], LSTM [322], GRU [323], CNN-1D [217], CNN-2D [324], CNN-3D [325], and MorphCNN [195] on the disjoint train–test dataset for the UP scene.

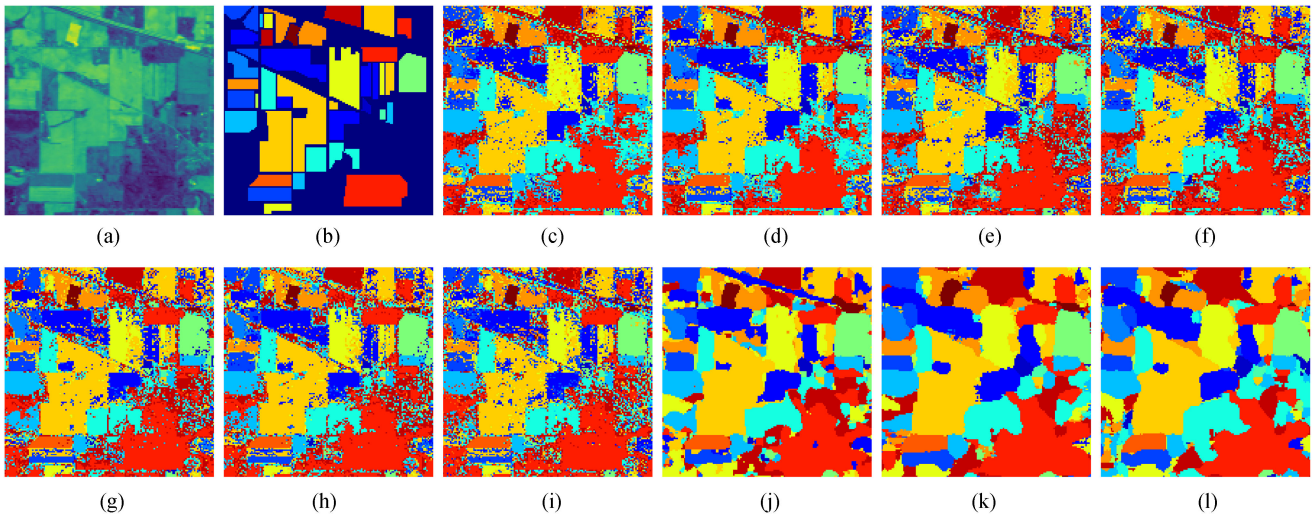


Fig. 21. Classification Maps obtained by MLR [320], SVM [321], MLP [69], RNN [70], LSTM [322], GRU [323], CNN-1D [217], CNN-2D [324], CNN-3D [325] and MorphCNN [195] on the disjoint train-test dataset for the IP scene.

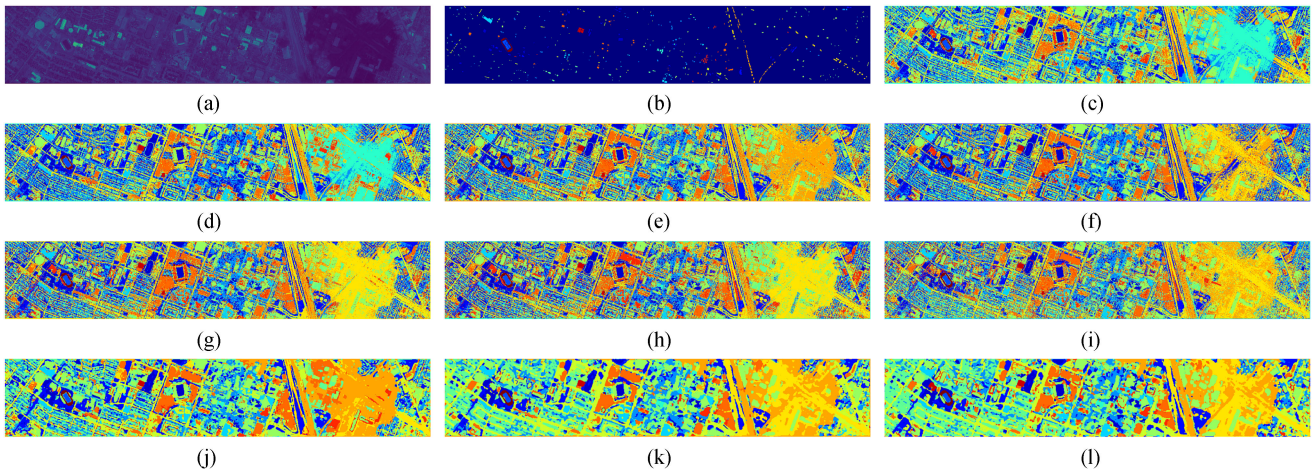


Fig. 22. Classification maps obtained by MLR [320], SVM [321], MLP [69], RNN [70], LSTM [322], GRU [323], CNN-1D [217], CNN-2D [324], CNN-3D [325], and MorphCNN [195] on the disjoint train–test dataset for the UH scene.

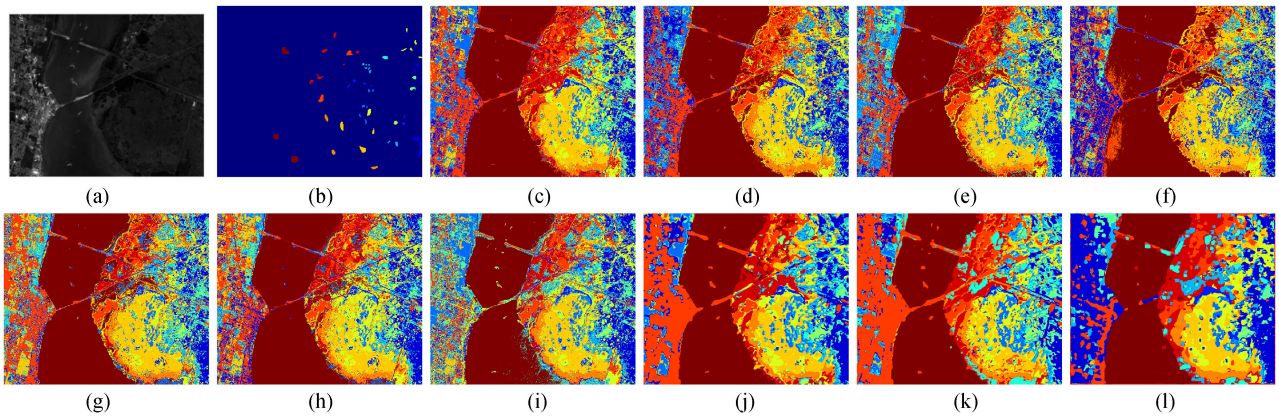


Fig. 23. Classification maps obtained by MLR [320], SVM [321], MLP [69], RNN [70], LSTM [322], GRU [323], CNN-1D [217], CNN-2D [324], CNN-3D [325], and MorphCNN [195] on the disjoint train–test dataset for the KSC scene.

TABLE VII
CLASSIFICATION RESULTS OBTAINED BY RF [75], MLR [320], SVM [321], MLP [69], RNN [70], LSTM [322], GRU [323], CNN-1D [217], CNN-2D [324], CNN-3D [325], HYBRIDSN [212], AND MORPHCNN [195] ON THE DISJOINT TRAIN–TEST DATASET FOR THE UNIVERSITY OF TRENTO (UT) SCENE

Class	RF [75]	MLR [321]	SVM [322]	MLP [69]	RNN [70]	LSTM [323]	GRU [324]	CNN-1D [218]	CNN-2D [325]	CNN-3D [326]	HybridSN [213]	MorphCNN [196]
1	97.27±0.33	92.57±0.00	95.03±0.00	70.68±3.69	93.46±3.67	95.71±0.32	95.78±0.97	98.06±0.72	96.35±0.22	92.64±7.45	**99.15±0.16**	97.83±0.37
2	89.50±0.11	**90.92±0.00**	88.66±0.00	76.22±2.90	83.09±1.50	84.73±0.87	84.40±1.51	89.16±1.99	85.87±1.87	75.70±7.23	81.65±2.93	90.41±1.83
3	75.04±0.33	90.10±0.00	91.71±0.00	83.95±2.08	69.51±0.87	59.26±14.31	65.15±4.84	70.67±8.99	66.84±2.00	60.07±3.57	74.86±10.58	83.86±10.69
4	**99.96±0.01**	92.77±0.00	85.18±0.00	37.55±2.74	98.44±0.97	97.72±1.41	99.65±0.18	99.86±0.11	98.67±0.61	98.88±0.80	99.64±0.28	99.06±0.57
5	**99.97±0.00**	99.14±0.00	97.76±0.00	99.98±.01	98.12±0.56	96.34±0.80	98.20±0.54	99.78±0.19	99.32±0.39	98.40±1.77	98.12±1.79	99.87±0.14
6	67.71±0.31	66.80±0.00	72.37±0.00	73.91±2.88	65.54±0.90	69.04±2.29	63.89±1.02	77.94±3.83	74.34±6.48	69.42±3.43	70.72±4.74	**82.64±0.88**
OA	94.95±0.07	92.08±0.00	89.98±0.00	63.51±9.43	92.43±0.34	92.27±0.72	93.03±0.28	95.94±0.23	94.45±0.13	92.14±1.05	94.03±0.21	**96.46±0.27**
AA	88.24±0.07	88.72±0.00	88.45±0.00	63.22±5.90	84.69±0.51	83.80±1.91	84.51±0.85	89.25±1.34	86.90±0.53	82.52±4.27	87.36±1.83	**92.28±1.60**
k(x100)	93.23±0.09	89.43±0.00	86.74±0.00	48.58±13.70	89.86±0.48	89.67±0.95	90.66±0.37	94.55±0.31	92.56±0.18	89.45±1.43	92.00±0.27	**95.26±0.36**

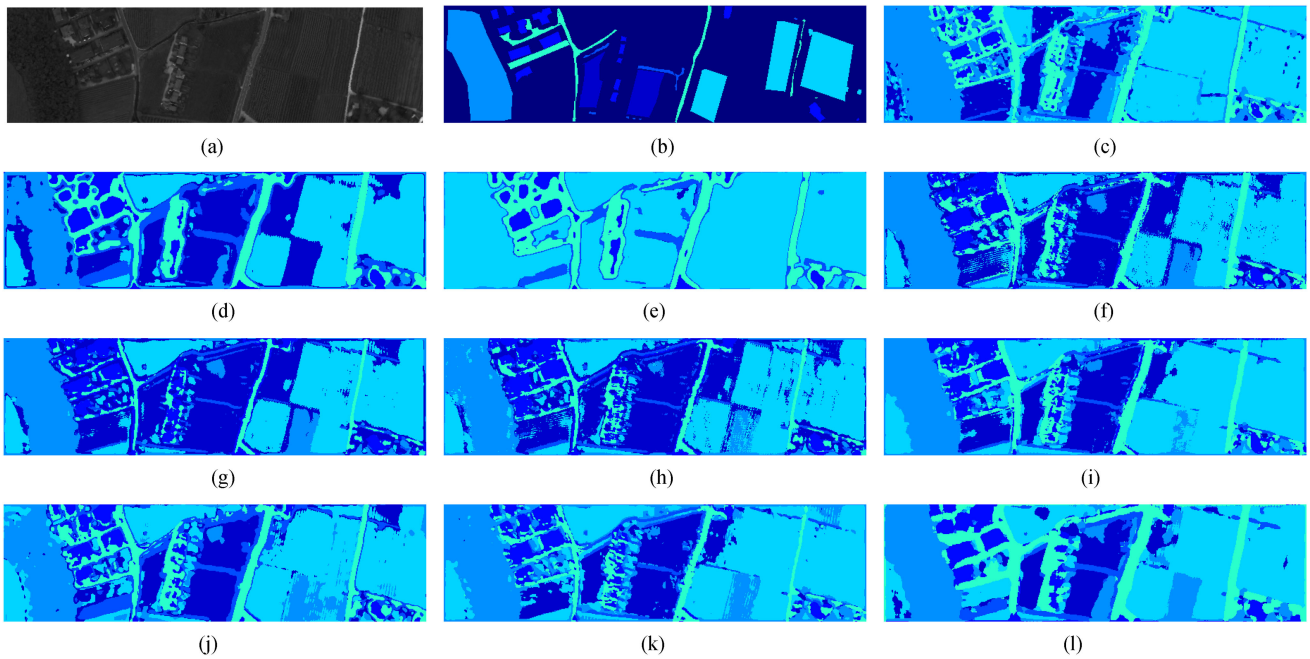


Fig. 24. Classification Maps obtained by MLR [320], SVM [321], MLP [69], RNN [70], LSTM [322], GRU [323], CNN-1D [217], CNN-2D [324], CNN-3D [325], and MorphCNN [195] on the disjoint train–test dataset for the UT scene.

by increasing the number of labeled training samples. Thus, a higher number of labeled training samples can produce better accuracies for all competing methods.

In general, the works [195], [324] outperformed (i.e., stable results) than the other comparative methods, especially in the case of less number of labeled training samples. The above leads to conclude that these works are not sensitive to the number of training samples. Moreover, as the number of training samples increases, the accuracies also increase for these methods; however, other methods can work better with a higher number of training samples as compared to these methods. A similar trend has been observed with a higher number of training samples. Thus, one can conclude that the works [195] and [324] can solve the limited availability of training samples issues to some extent while considering disjoint train/test samples.

Moreover, one can conclude that the AE-based models do not perform well as compared to the other models, although the unsupervised methods do not require the samples to be labeled; if there are no constraints, these methods might learn nothing. Moreover, AE has a symmetric architecture that leads to the explosion of training parameters, which increases the difficulty in training. The works [328] and [329] overcome the abovementioned issues; however, the work [227] does not adopt the greedy layer-wise approach, thus producing the worst results; thus, there is room for improvement in such methods.

In a nutshell, the classification results based on CNN are way better than AE-based methods while considering the limited availability of labeled training samples. Although the AEs can learn the internal structure of the unlabeled data, the final feature representation might not have task-driven characteristics, which might be the reason for not performing well as compared to the supervised learning models. Moreover, AL and/or SL takes the benefits from the selection of the most important samples for training, which enables the model to focus more attention on indistinguishable samples for HSIC, whereas few-shot learning (FSL) benefits from the exploration of the relationship between samples to find a discriminative decision boundary for HSIC. TL makes good use of similarity among different HSIs to reduce the quantity required for training and also reduces the number of trainable parameters while boosting the models' robustness. According to the raw data (i.e., processing the HSI without extracting/learning the features), DA generates more samples which bring a diversity of samples.

TABLE VIII
CLASSIFICATION RESULTS OBTAINED BY CNN-2D [324], CNN-3D [325], G2C-CONV2D [215], AND G2C-CONV3D [215] ON THE DISJOINT TRAIN–TEST FOR IP, PU, TRENTO, UH, AND KSC DATASETS. THE HIGHER ACCURACIES ARE EMPHAZISED

Class	IP				PU				UH				Trento				KSC			
	Conv2D	Conv3D	G2C-2D	G2C-3D	Conv2D	Conv3D	G2C-2D	G2C-3D	Conv2D	Conv3D	G2C-2D	G2C-3D	Conv2D	Conv3D	G2C-2D	G2C-3D	Conv2D	Conv3D	G2C-2D	G2C-3D
1	46.66±0.09	56.00±8.64	53.33±18.57	**97.33±1.88**	81.20±0.01	87.20±0.42	86.22±4.54	**91.26±3.40**	80.88±0.006	82.77±0.27	81.32±0.08	**82.90±0.20**	98.30±0.007	98.63±0.18	**98.75±0.58**	96.12±1.95	95.56±0.007	98.40±0.81	98.24±0.52	**99.27±0.56**
2	48.04±0.03	64.64±4.75	**68.93±5.41**	57.33±4.91	89.52±0.01	89.90±0.35	**92.99±0.90**	83.84±0.49	81.64±0.005	82.64±0.85	81.79±1.44	**83.52±0.31**	85.74±0.029	84.78±2.61	90.29±2.68	**93.30±1.41**	79.06±0.03	95.16±1.18	90.82±2.04	**98.87±0.60**
3	24.33±0.009	44.96±8.99	35.97±6.92	**78.21±2.13**	53.53±0.006	59.98±2.20	51.33±2.60	**78.08±2.84**	49.37±0.032	65.80±5.12	74.32±7.73	**91.48±1.12**	78.43±0.08	**86.63±9.87**	63.81±7.90	83.51±12.16	74.92±0.016	90.36±5.99	80.58±3.37	**97.09±0.21**
4	28.28±0.13	41.75±4.15	18.18±2.18	**42.42±3.59**	96.92±0.009	**97.43±0.44**	96.97±1.05	94.05±1.16	86.26±0.006	**95.54±3.17**	91.60±0.29	90.49±2.39	93.59±0.006	98.31±0.30	97.11±0.76	**98.53±0.23**	50.62±0.056	74.61±3.33	69.62±4.01	**77.57±1.52**
5	29.56±0.05	46.83±2.99	26.88±3.96	**52.79±3.88**	98.41±0.007	97.99±0.95	**99.28±0.38**	99.19±0.38	95.39±0.018	96.96±0.27	98.67±1.00	**99.87±0.08**	98.22±0.011	**99.77±0.18**	99.57±0.22	98.70±0.60	71.04±0.03	**88.32±3.31**	81.26±1.37	87.34±0.91
6	72.31±0.10	92.56±3.01	94.35±2.59	**98.39±1.06**	47.65±0.08	62.30±2.37	60.68±4.23	**87.04±2.05**	81.11±0.02	92.07±1.83	82.51±4.67	**94.40±0.00**	65.09±0.036	75.74±3.30	58.87±2.95	73.78±2.48	79.82±0.03	90.25±1.10	78.63±4.43	**93.84±1.10**
7	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	61.36±0.02	76.55±1.11	65.30±3.66	**88.44±5.08**	79.57±0.02	84.48±2.21	82.64±1.12	**85.85±0.49**					88.01±0.010	98.50±1.40	95.88±1.05	**99.62±0.52**
8	91.60±0.05	**98.93±0.67**	98.66±0.82	94.53±4.64	81.97±0.02	94.73±1.51	97.25±0.93	**97.34±0.92**	47.61±0.08	65.71±4.44	56.98±0.73	**71.06±0.98**					85.15±0.062	97.08±0.34	87.70±1.77	**98.99±0.71**
9	36.66±0.04	33.33±12.47	50.00±8.16	**80.00±8.16**	88.38±0.04	**96.98±0.81**	96.31±1.74	96.85±1.61	74.15±0.04	79.85±1.96	77.49±1.45	**82.37±1.60**					98.11±0.013	98.64±0.97	99.62±0.21	**99.77±0.18**
10	50.49±0.04	**62.42±5.93**	52.75±3.42	55.00±4.17					41.40±0.02	45.65±1.29	**52.67±3.07**	49.25±1.74					96.20±0.008	99.31±0.76	99.90±0.13	**100.0±0.00**
11	73.45±0.01	79.68±1.91	79.68±0.66	**81.72±0.19**					50.18±0.01	54.26±3.88	65.71±1.04	**70.55±1.47**					92.97±0.007	99.81±0.26	99.15±0.45	**100.0±0.00**
12	30.61±0.01	49.29±4.92	**56.38±5.87**	35.93±7.39					70.22±0.02	76.46±3.88	**88.98±4.50**	88.05±4.16					92.83±0.025	95.63±0.77	**98.90±0.39**	98.13±0.66
13	95.83±0.03	**97.08±1.17**	94.16±5.62	95.83±0.58					85.84±0.019	**91.92±0.28**	85.02±0.92	82.57±0.82					**100.0±0.00**	**100.0±0.00**	**100.0±0.00**	**100.0±0.00**
14	74.18±0.001	76.08±4.46	**95.59±0.98**	90.03±5.77					77.86±0.06	82.86±1.63	80.70±0.68	**99.05±0.76**								
15	19.52±0.01	54.54±16.05	21.21±8.12	**74.74±2.97**					48.34±0.06	70.68±3.24	59.83±6.72	**99.92±0.09**								
16	71.21±0.18	**78.78±2.83**	62.87±5.96	78.03±10.55																
OA	57.02±0.12	69.25±1.32	68.33±0.60	**72.82±0.26**	81.23±0.36	85.96±0.27	86.55±0.44	**87.79±0.23**	69.67±0.19	76.52±1.13	77.27±0.61	**82.26±0.56**	91.95±1.03	95.09±0.26	93.15±0.34	**95.21±0.33**	89.76±1.13	96.15±0.43	94.05±0.49	**97.65±0.02**
AA	49.55±0.01	61.05±2.92	56.81±1.73	**69.52±1.23**	77.66±0.002	84.78±0.18	82.92±0.38	**90.68±0.06**	69.99±0.007	77.84±0.86	77.35±1.18	**84.76±0.52**	86.56±0.021	90.64±1.80	84.73±1.57	**90.66±2.24**	84.95±0.011	94.31±0.65	90.79±0.64	**96.19±0.01**
k(x100)	50.66±0.22	64.78±1.58	63.60±0.74	**68.87±0.37**	74.57±0.58	81.11±0.36	81.74±0.65	**83.95±0.30**	67.22±0.24	74.61±1.19	75.43±0.67	**80.87±0.60**	89.20±1.38	93.42±0.36	90.79±0.46	**93.62±0.45**	88.60±1.26	95.72±0.48	93.37±0.55	**97.39±0.02**

C. Experiments With Convolutional Feature Extractors

This section revisited several deep hyperspectral FE processes, i.e., a traditional convolutional process and a gradient centralized convolutional process. In this hierarchy, we have conducted several experiments using several state-of-the-art works published in recent years. This experiment is specifically designed to check the performance of the convolutional process rather than testing the model's performance. The baseline models apply convolutional feature extractors which include a 2-D convolution neural network for HSIC (Conv2D) introduced by Makantasis *et al.* [324] and the 3-D convolutional approach for remote sensing image classification (Conv3D) proposed by Hamida *et al.* [325] (a traditional 3-D convolutional feature extractor), and, recently, Roy *et al.* [215] introduced generalized gradient centralized 2-D convolution (G2C-Conv2D), and generalized gradient centralized 3-D convolution (G2C-Conv3D) [215] to extract the fine-grained spectral–spatial feature representation. The generalized gradient centralized 3-D convolution (G2C-Conv3D) operation is designed by using a weighted combination between the vanilla and gradient centralized 3-D convolutions (GC-Conv3D) to extract both the intensity-level semantic information and gradient-level information from the HSIs.

All the aforementioned convolutional feature extractors have been evaluated on five different hyperspectral datasets, namely, IP, PU, Trento, UH, and KSC datasets. The experimental results are illustrated in Table VIII. From all these results, one can easily conclude that the G2C-Conv3D convolutional process outperformed Conv2D and Conv3D, followed by G2C-Conv2D. A similar trend has been observed for all datasets except the Trento dataset on which the 3-D convolutional process slightly performed better as compared to the traditional Conv2D and G2C-Conv2D, respectively. The accuracy difference is not that high as compared to the G2C-Conv3D for other datasets. Most importantly, the G2C-Conv3D convolution operation is simple to implement and can easily be plugged into existing CNNs to boost both the robustness and classification performance.

XII. CONCLUSION

The rich information contained in HSI data is a captivating factor that constitutes the utilization of HSI technology in real-world applications. Moreover, advances in machine learning methods strengthen the deployment potentials of such technologies. In this article, we surveyed recent developments of HSIC using state-of-the-art DNNs [for instance, AE, DBN, RNN, CNN, TL, FSL, active/self learning (AL/SL), and DA] in a variety of learning schemes (specifically, supervised, semi-supervised, and unsupervised learning). In addition, we also analyzed the strategies to overcome the challenges of limited availability of training data like DA, FSL, TL, and AL. According to the methodologies discussed above, we select some of the representative works to conduct the experiments on benchmark HSI datasets.

Although the current HSIC techniques reflect a rapid, remarkable, and sophistication of the task, further developments are still required to improve the generalization capabilities. The main issue of DNN-based HSIC is the lack of labeled data. HSI data is infamous due to the limited availability of labeled data and DNNs demand a sufficiently large amount of labeled training data. Section X discussed some widely used strategies to combat the aforesaid issue but significant improvements are still needed to efficiently utilize limited available training data. One direction to solve this problem could be to explore the integration of various learning strategies discussed in Section X to cash in the joint benefits. One more way is to exploit a few-shot or K-shot learning approaches that can accurately predict the class labels with only a few labeled samples. Moreover, there is a need to focus on the joint exploitation of spectral–spatial features of HSI to complement classification accuracies achieved from the aforementioned HSIC frameworks. Another future potential of HSIC is computationally efficient architectures. Therefore, the issue of the high computational complexity of DNNs is of paramount importance and it is crucial to implement parallel HSIC architectures to speed up the processing of DNNs to meet the computational stipulation of time-critical HSI applications. In this direction, HPC platforms and specialized hardware modules like GPUs and FPGAs can be used to implement the parallel HSIC frameworks. Hence, to assimilate the aforesaid aspects in the development of a new HSIC framework is to appropriately utilize the limited training samples while considering joint spectral–spatial features of HSI and maintaining the low computational burden.

ACKNOWLEDGMENT

The authors thank Ganesan Narayanasamy who is leading IBM OpenPOWER/POWER enablement and ecosystem worldwide for his support to get the IBM AC922 system's access.

REFERENCES

[1] M. Ahmad *et al.*, "Spatial prior fuzziness pool-based interactive classification of hyperspectral images," *Remote Sens.*, vol. 11, no. 9, pp. 1–19. May 2019.

[2] D. Hong *et al.*, "Interpretable hyperspectral artificial intelligence: When non-convex modeling meets hyperspectral remote sensing," *IEEE Geosci. Remote Sens. Mag.*, vol. 9, no. 2, pp. 52–87, Jun. 2021.

[3] H. Ayaz *et al.*, "Myoglobin-based classification of minced meat using hyperspectral imaging," *Appl. Sci.*, vol. 10, no. 19, p. 6862, 2020.

[4] M. H. Khan, Z. Saleem, M. Ahmad, A. Sohaib, H. Ayaz, and M. Mazzara, "Hyperspectral imaging for color adulteration detection in red chili," *Appl. Sci.*, vol. 10, no. 17, p. 5955, 2020.

[5] Z. Saleem, M. H. Khan, M. Ahmad, A. Sohaib, H. Ayaz, and M. Mazzara, "Prediction of microbial spoilage and shelf-life of bakery products through hyperspectral imaging," *IEEE Access*, vol. 8, pp. 176986–176996, 2020.

[6] M. Zulfiqar, M. Ahmad, A. Sohaib, M. Mazzara, and S. Distefano, "Hyperspectral imaging for bloodstain identification," *Sensors*, vol. 21, no. 9, p. 3045, 2021.

[7] H. Ayaz, M. Ahmad, M. Mazzara, and A. Sohaib, "Hyperspectral imaging for minced meat classification using nonlinear deep features," *Appl. Sci.*, vol. 10, no. 21, p. 7783, 2020.

[8] H. Hussain Khan *et al.*, "Hyperspectral imaging-based unsupervised adulterated red chili content transformation for classification: Identification of red chili adulterants," *Neural Comput. Appl.*, vol. 33, pp. 1–15, 2021.

[9] "10 Important applications of hyperspectral image," Accessed on: Mar. 10, 2020, [Online]. Available: https://grindgis.com/remote-sensing/10-important-applications-of-hyperspectral-image

[10] F. Xing, H. Yao, Y. Liu, X. Dai, R. L. Brown, and D. Bhatnagar, "Recent developments and applications of hyperspectral imaging for rapid detection of mycotoxins and mycotoxigenic fungi in food products," *Crit. Rev. Food Sci. Nutr.*, vol. 59, no. 1, pp. 173–180, 2019.

[11] M. Ahmad, "Ground truth labeling and samples selection for hyperspectral image classification," *Optik*, vol. 230, 2021, Art. no. 166267.

[12] "Applications of hyperspectral image," Accessed: Mar. 10, 2020. [Online]. Available: https://resonon.com/applications

[13] M. Ahmad, A. M. Khan, and R. Hussain, "Graph-based spatial spectral feature learning for hyperspectral image classification," *IET Image Process.*, vol. 11, no. 12, pp. 1310–1316, 2017.

[14] J. M. Haut, M. E. Paoletti, J. Plaza, and A. Plaza, "Fast dimensionality reduction and classification of hyperspectral images with extreme learning machines," *J. Real-Time Image Process.*, vol. 15, no. 3, pp. 439–462, 2018.

[15] D. Hong, N. Yokoya, J. Chanussot, J. Xu, and X. Zhu, "Learning to propagate labels on graphs: An iterative multitask regression framework for semi-supervised hyperspectral dimensionality reduction," *ISPRS J. Photogrammetry Remote Sens.*, vol. 158, pp. 35–49, 2019.

[16] M. Ahmad, S. Lee, D. Ulhaq, and Q. Mushtaq, "Hyperspectral remote sensing: Dimensional reduction and end member extraction," *Int. J. Soft Comput. Eng.*, vol. 2, pp. 2231–2307, 2012.

[17] D. Hong, N. Yokoya, J. Chanussot, J. Xu, and X. X. Zhu, "Joint and progressive subspace analysis (JPSA) with spatial–spectral manifold alignment for semi-supervised hyperspectral dimensionality reduction," *IEEE Trans. Cybern.*, vol. 51, no. 7, pp. 3602–3615, Jul. 2021.

[18] B. Zhang, X. Sun, L. Gao, and L. Yang, "Endmember extraction of hyperspectral remote sensing images based on the ant colony optimization (ACO) algorithm," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 7, pp. 2635–2646, Jul. 2011.

[19] J. M. Bioucas-Dias *et al.*, "Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 2, pp. 354–379, Apr. 2012.

[20] M. Ahmad, A. Khan, and A. K. Bashir, "Metric similarity regularizer to enhance pixel similarity performance for hyperspectral unmixing," *Optik – Int. J. Light Electron Opt.*, vol. 140, pp. 86–95, Jul. 2017.

[21] Y. Zhong, X. Wang, L. Zhao, R. Feng, L. Zhang, and Y. Xu, "Blind spectral unmixing based on sparse component analysis for hyperspectral remote sensing imagery," *ISPRS J. Photogrammetry Remote Sens.*, vol. 119, pp. 49–63, 2016.

[22] M. Ahmad, D. Ihsan, and D. Ulhaq, "Linear unmixing and target detection of hyperspectral imagery using OSP," in *Proc. Int. Proc. Comput. Sci. Inf. Technol.*, 2011, pp. 179–183.

[23] M. Ahmad, D. Ulhaq, Q. Mushtaq, and M. Sohaib, "A new statistical approach for band clustering and band selection using K-means clustering," *Int. J. Eng. Technol.*, vol. 3, pp. 606–614, 2011.

[24] M. Ahmad, D. Ulhaq, and Q. Mushtaq, "AIK method for band clustering using statistics of correlation and dispersion matrix," in *Proc. Int. Conf. Inf. Commun. Manage.*, 2011, pp. 114–118.

[25] D. Hong, N. Yokoya, J. Chanussot, and X. X. Zhu, "An augmented linear mixing model to address spectral variability for hyperspectral unmixing," *IEEE Trans. Image Process.*, vol. 28, no. 4, pp. 1923–1938, Apr. 2019.

[26] L. Gao, Z. Han, D. Hong, B. Zhang, and J. Chanussot, "CyCU-Net: Cycle-consistency unmixing network by learning cascaded autoencoders," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art no. 5503914, doi: 10.1109/TGRS.2021.3064958.

[27] D. W. Stein, S. G. Beaven, L. E. Hoff, E. M. Winter, A. P. Schaum, and A. D. Stocker, "Anomaly detection from hyperspectral imagery," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 58–69, Jan. 2002.

[28] S. Liu, Q. Du, X. Tong, A. Samat, L. Bruzzone, and F. Bovolo, "Multiscale morphological compressed change vector analysis for unsupervised multiple change detection," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 9, pp. 4124–4137, Sep. 2017.

[29] S. Li, K. Zhang, Q. Hao, P. Duan, and X. Kang, "Hyperspectral anomaly detection with multiscale attribute and edge-preserving filters," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 10, pp. 1605–1609, Oct. 2018.

[30] X. Wu, D. Hong, J. Tian, J. Chanussot, W. Li, and R. Tao, "ORSIm detector: A novel object detection framework in optical remote sensing imagery using spatial-frequency channel features," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 7, pp. 5146–5158, Jul. 2019.

[31] S. Liu, D. Marinelli, L. Bruzzone, and F. Bovolo, "A review of change detection in multitemporal hyperspectral images: Current techniques, applications, and challenges," *IEEE Geosci. Remote Sens. Mag.*, vol. 7, no. 2, pp. 140–158, Jun. 2019.

[32] X. Wu, D. Hong, J. Chanussot, Y. Xu, R. Tao, and Y. Wang, "Fourier-based rotation-invariant feature boosting: An efficient framework for geospatial object detection," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 2, pp. 302–306, Feb. 2020.

[33] P. Chen *et al.*, "FCCDN: Feature constraint network for VHR image change detection," 2021, *arXiv:2105.10860*.

[34] M. Fauvel, Y. Tarabalka, J. A. Benediktsson, J. Chanussot, and J. C. Tilton, "Advances in spectral-spatial classification of hyperspectral images," in *Proc. IEEE*, vol. 101, no. 3, pp. 652–675, Mar. 2013.

[35] M. Ahmad, S. Protasov, A. M. Khan, R. Hussain, A. M. Khattak, and W. A. Khan, "Fuzziness-based active learning framework to enhance hyperspectral image classification performance for discriminative and generative classifiers," *PLoS ONE*, vol. 13, Jan. 2018, Art. no. 188996.

[36] M. Ahmad, S. Shabbir, D. Oliva, M. Mazzara, and S. Distefano, "Spatial-prior generalized fuzziness extreme learning machine autoencoder-based active learning for hyperspectral image classification," *Optik – Int. J. Light Electron Opt.*, 2020, Art. no. 163712.

[37] M. Ahmad, A. M. Khan, R. Hussain, S. Protasov, F. Chow, and A. M. Khattak, "Unsupervised geometrical feature learning from hyperspectral data," in *Proc. IEEE Symp. Ser. Comput. Intell.*, 2016, pp. 1–6.

[38] M. Ahmad, S. Protasov, and A. M. Khan, "Hyperspectral band selection using unsupervised non-linear deep auto encoder to train external classifiers," *CoRR*, 2017, *arXiv: 1705.06920*.

[39] M. Ahmad, M. A. Alqarni, A. M. Khan, R. Hussain, M. Mazzara, and S. Distefano, "Segmented and non-segmented stacked denoising autoencoder for hyperspectral band reduction," *Optik – Int. J. Light Electron Opt.*, vol. 180, pp. 370–378, Oct. 2018.

[40] S. Liu, Y. Zheng, Q. Du, A. Samat, X. Tong, and M. Dalponte, "A novel feature fusion approach for VHR remote sensing image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 464–473, 2020.

[41] D. Hong, N. Yokoya, J. Chanussot, and X. X. Zhu, "CoSpace: Common subspace learning from hyperspectral-multispectral correspondences," *IEEE Trans. Geos. Remote Sens.*, vol. 57, no. 7, pp. 4349–4359, Jul. 2019.

[42] L. Gao, Q. Du, B. Zhang, W. Yang, and Y. Wu, "A comparative study on linear regression-based noise estimation for hyperspectral imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 6, no. 2, pp. 488–498, Apr. 2013.

[43] W. Wei, L. Zhang, C. Tian, A. Plaza, and Y. Zhang, "Structured sparse coding-based hyperspectral imagery denoising with intracluster filtering," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 12, pp. 6860–6876, Dec. 2017.

[44] C. Yi, Y.-Q. Zhao, J. Yang, J. C.-W. Chan, and S. G. Kong, "Joint hyperspectral superresolution and unmixing with interactive feedback," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 7, pp. 3823–3834, Jul. 2017.

[45] C. Yi, Y.-Q. Zhao, and J. C.-W. Chan, "Hyperspectral image superresolution based on spatial and spectral correlation fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 7, pp. 4165–4177, Jul. 2018.

[46] G. Cheng, J. Han, L. Guo, Z. Liu, S. Bu, and J. Ren, "Effective and efficient midlevel visual elements-oriented land-use classification using VHR remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 8, pp. 4238–4249, Aug. 2015.

[47] Q. Zhu, Y. Zhong, B. Zhao, G.-S. Xia, and L. Zhang, "Bag-of-visual-words scene classifier with local and global features for high spatial resolution remote sensing imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 6, pp. 747–751, Jun. 2016.

[48] H. Wu, B. Liu, W. Su, W. Zhang, and J. Sun, "Hierarchical coding vectors for scene level land-use classification," *Remote Sens.*, vol. 8, no. 5, p. 436, 2016.

[49] G. Cheng, J. Han, and X. Lu, "Remote sensing image scene classification: Benchmark and state of the art," in *Proc. IEEE*, vol. 105, no. 10, pp. 1865–1883, Oct. 2017.

[50] D. Hong, N. Yokoya, N. Ge, J. Chanussot, and X. Zhu, "Learnable manifold alignment (LeMA): A semi-supervised cross-modality learning framework for land cover and land use classification," *ISPRS J. Photogrammetry Remote Sens.*, vol. 147, pp. 193–205, 2019.

[51] T. R. Martha, N. Kerle, C. J. van Westen, V. Jetten, and K. V. Kumar, "Segment optimization and data-driven thresholding for knowledge-based landslide detection by object-based image analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 12, pp. 4928–4943, Dec. 2011.

[52] G. Cheng, L. Guo, T. Zhao, J. Han, H. Li, and J. Fang, "Automatic landslide detection from remote-sensing imagery using a scene classification method based on BoVW and pLSA," *Int. J. Remote Sens.*, vol. 34, no. 1, pp. 45–59, 2013.

[53] N. B. Mishra and K. A. Crews, "Mapping vegetation morphology types in a dry savanna ecosystem: Integrating hierarchical object-based image analysis with random forest," *Int. J. Remote Sens.*, vol. 35, no. 3, pp. 1175–1198, 2014.

[54] X. Li and G. Shao, "Object-based urban vegetation mapping with high-resolution aerial photography as a single data source," *Int. J. Remote Sens.*, vol. 34, no. 3, pp. 771–789, 2013.

[55] S. B. Kotsiantis, I. D. Zaharakis, and P. E. Pintelas, "Machine learning: A review of classification and combining techniques," *Artif. Intell. Rev.*, vol. 26, no. 3, pp. 159–190, 2006.

[56] S. B. Kotsiantis, I. Zaharakis, and P. Pintelas, "Supervised machine learning: A review of classification techniques," *Emerg. Artif. Intell. Appl. Comput. Eng.*, vol. 160, pp. 3–24, 2007.

[57] C. Pan, X. Jia, J. Li, and X. Gao, "Adaptive edge preserving maps in Markov random fields for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 10, pp. 8568–8583, Oct. 2021.

[58] A. Plaza *et al.*, "Recent advances in techniques for hyperspectral image processing," *Remote Sens. Environ.*, vol. 113, pp. S110–S122, 2009.

[59] R. Ablin and C. H. Sulochana, "A survey of hyperspectral image classification in remote sensing," *Int. J. Adv. Res. Comput. Commun. Eng.*, vol. 2, no. 8, pp. 2986–3000, 2013.

[60] 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.

[61] D. Chutia, D. Bhattacharyya, K. K. Sarma, R. Kalita, and S. Sudhakar, "Hyperspectral remote sensing classifications: A perspective survey," *Trans. GIS*, vol. 20, no. 4, pp. 463–490, 2016.

[62] Y. Chen, Z. 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.

[63] P. Ghamisi *et al.*, "Advances in hyperspectral image and signal processing: A comprehensive overview of the state of the art," *IEEE Geosci. Remote Sens. Mag.*, vol. 5, no. 4, pp. 37–78, Dec. 2017.

[64] C. Li, Y. Wang, X. Zhang, H. Gao, Y. Yang, and J. Wang, "Deep belief network for spectral–spatial classification of hyperspectral remote sensor data," *Sensors*, vol. 19, no. 1, p. 204, 2019.

[65] B. Rasti *et al.*, "Feature extraction for hyperspectral imagery: The evolution from shallow to deep: Overview and toolbox," *IEEE Geosci. Remote Sens. Mag.*, vol. 8, no. 4, pp. 60–88, Dec. 2020.

[66] H. Petersson, D. Gustafsson, and D. Bergstrom, "Hyperspectral image analysis using deep learning-A review," in *Proc. 6th Int. Conf. Image Process. Theory, Tools Appl.*, 2016, pp. 1–6.

[67] M. Ahmad, A. M. Khan, M. Mazzara, S. Distefano, M. Ali, and M. S. Sarfraz, "A fast and compact 3-D CNN for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, to be published, doi: 10.1109/LGRS.2020.3043710.

[68] D. Hong *et al.*, "SpectralFormer: Rethinking hyperspectral image classification with transformers," *IEEE Trans. Geosci. Remote Sens*., to be published, doi: 10.1109/TGRS.2021.3130716.

[69] M. Paoletti, J. Haut, J. Plaza, and A. Plaza, "Deep learning classifiers for hyperspectral imaging: A review," *ISPRS J. Photogrammetry Remote Sens.*, vol. 158, pp. 279–317, 2019.

[70] R. Hang, Q. Liu, D. Hong, and P. Ghamisi, "Cascaded recurrent neural networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 8, pp. 5384–5394, Aug. 2019.

[71] L. Huang, C. Chen, W. Li, and Q. Du, "Remote sensing image scene classification using multi-scale completed local binary patterns and fisher vectors," *Remote Sens.*, vol. 8, no. 6, p. 483, 2016.

[72] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2005, vol. 1, pp. 886–893.

[73] A. Oliva and A. Torralba, "Modeling the shape of the scene: A holistic representation of the spatial envelope," *Int. J. Comput. Vis.*, vol. 42, no. 3, pp. 145–175, 2001.

[74] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proc. 7th IEEE Int. Conf. Comput. Vis.*, 1999, vol. 2, pp. 1150–1157.

[75] J. Ham, Y. Chen, M. M. Crawford, and J. Ghosh, "Investigation of the random forest framework for classification of hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 492–501, Mar. 2005.

[76] G. Camps-Valls and L. Bruzzone, "Kernel-based methods for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 6, pp. 1351–1362, Jun. 2005.

[77] W. Yang, L. Gao, and D. Chen, "Real-time target detection in hyperspectral images based on spatial-spectral information extraction," *EURASIP J. Adv. Signal Process.*, vol. 2012, pp. 1–15, 2012.

[78] G. Cheng, P. Zhou, X. Yao, C. Yao, Y. Zhang, and J. Han, "Object detection in VHR optical remote sensing images via learning rotation-invariant hog feature," in *Proc. 4th Int. Workshop Earth Observ. Remote Sens. Appl.*, 2016, pp. 433–436.

[79] G. Cheng, P. Zhou, J. Han, L. Guo, and J. Han, "Auto-encoder-based shared mid-level visual dictionary learning for scene classification using very high resolution remote sensing images," *IET Comput. Vis.*, vol. 9, no. 5, pp. 639–647, 2015.

[80] R. Azhar *et al.*, "Batik image classification using SIFT feature extraction, bag of features and support vector machine," *Procedia Comput. Sci.*, vol. 72, pp. 24–30, 2015.

[81] O. Zeglazi, A. Amine, and M. Rziza, "SIFT descriptors modeling and application in texture image classification," in *Proc. 13th Int. Conf. Comput. Graph., Imag. Visualization*, 2016, pp. 265–268.

[82] Y. Xu, K. Hu, Y. Tian, and F. Peng, "Classification of hyperspectral imagery using SIFT for spectral matching," in *Proc. Congr. Image Signal Process.*, 2008, pp. 704–708.

[83] Y. Yang and S. Newsam, "Comparing SIFT descriptors and Gabor texture features for classification of remote sensed imagery," in *Proc. 15th IEEE Int. Conf. Image Process.*, 2008, pp. 1852–1855.

[84] H. T. M. Nhat and V. T. Hoang, "Feature fusion by using LBP, HOG, GIST descriptors and canonical correlation analysis for face recognition," in *Proc. 26th Int. Conf. Telecommun.*, 2019, pp. 371–375.

[85] S. K. Roy, B. Chanda, B. B. Chaudhuri, D. K. Ghosh, and S. R. Dubey, "Local morphological pattern: A scale space shape descriptor for texture classification," *Digit. Signal Process.*, vol. 82, pp. 152–165, 2018.

[86] R. M. Haralick, K. Shanmugam, and I. H. Dinstein, "Textural features for image classification," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-3, no. 6, pp. 610–621, Nov. 1973.

[87] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971–987, Jul. 2002.

[88] L. Zhao, P. Tang, and L. Huo, "Feature significance-based multibag-of-visual-words model for remote sensing image scene classification," *J. Appl. Remote Sens.*, vol. 10, no. 3, 2016, Art.no. 035004.

[89] Y. Zhang, X. Sun, H. Wang, and K. Fu, "High-resolution remote-sensing image classification via an approximate Earth mover's distance-based bag-of-features model," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 5, pp. 1055–1059, Sep. 2013.

[90] Y. Yang and S. Newsam, "Bag-of-visual-words and spatial extensions for land-use classification," in *Proc. 18th SIGSPATIAL Int. Conf. Adv. Geographic Inf. Syst.*, 2010, pp. 270–279.

[91] S. Xu, T. Fang, D. Li, and S. Wang, "Object classification of aerial images with bag-of-visual words," *IEEE Geosci. Remote Sens. Lett.*, vol. 7, no. 2, pp. 366–370, Apr. 2010.

[92] J. Zhang, T. Li, X. Lu, and Z. Cheng, "Semantic classification of high-resolution remote-sensing images based on mid-level features," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 6, pp. 2343–2353, Jun. 2016.

[93] R. Bahmanyar, S. Cui, and M. Datcu, "A comparative study of bag-of-words and bag-of-topics models of EO image patches," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 6, pp. 1357–1361, Jun. 2015.

[94] G. Cheng, J. Han, P. Zhou, and L. Guo, "Multi-class geospatial object detection and geographic image classification based on collection of part detectors," *ISPRS J. Photogrammetry Remote Sens.*, vol. 98, pp. 119–132, 2014.

[95] B. Zhao, Y. Zhong, L. Zhang, and B. Huang, "The Fisher kernel coding framework for high spatial resolution scene classification," *Remote Sens.*, vol. 8, no. 2, p. 157, 2016.

[96] J. Hu, G.-S. Xia, F. Hu, and L. Zhang, "A comparative study of sampling analysis in the scene classification of optical high-spatial resolution remote sensing imagery," *Remote Sens.*, vol. 7, no. 11, pp. 14988–15013, 2015.

[97] S. Lazebnik, C. Schmid, and J. Ponce, "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2006, vol. 2, pp. 2169–2178.

[98] B. Zhao, Y. Zhong, G.-S. Xia, and L. Zhang, "Dirichlet-derived multiple topic scene classification model for high spatial resolution remote sensing imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 4, pp. 2108–2123, Apr. 2016.

[99] R. Kusumaningrum, H. Wei, R. Manurung, and A. Murni, "Integrated visual vocabulary in latent Dirichlet allocation-based scene classification for IKONOS image," *J. Appl. Remote Sens.*, vol. 8, no. 1, 2014, Art. no. 0 83690.

[100] Y. Zhong, Q. Zhu, and L. Zhang, "Scene classification based on the multifeature fusion probabilistic topic model for high spatial resolution remote sensing imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 11, pp. 6207–6222, Nov. 2015.

[101] D. Hong, N. Yokoya, G.-S. Xia, J. Chanussot, and X. X. Zhu, "X-ModalNet: A semi-supervised deep cross-modal network for classification of remote sensing data," *ISPRS J. Photogrammetry Remote Sens.*, vol. 167, pp. 12–23, 2020.

[102] H. Yu, W. Yang, G.-S. Xia, and G. Liu, "A color-texture-structure descriptor for high-resolution satellite image classification," *Remote Sens.*, vol. 8, no. 3, p. 259, 2016.

[103] V. Risojević and Z. Babić, "Fusion of global and local descriptors for remote sensing image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 4, pp. 836–840, Jul. 2013.

[104] M. L. Mekhalfi, F. Melgani, Y. Bazi, and N. Alajlan, "Land-use classification with compressive sensing multifeature fusion," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 10, pp. 2155–2159, Oct. 2015.

[105] G. Sheng, W. Yang, T. Xu, and H. Sun, "High-resolution satellite scene classification using a sparse coding based multiple feature combination," *Int. J. Remote Sens.*, vol. 33, no. 8, pp. 2395–2412, 2012.

[106] L. Xie, J. Wang, B. Zhang, and Q. Tian, "Incorporating visual adjectives for image classification," *Neurocomputing*, vol. 182, pp. 48–55, 2016.

[107] D. Hong *et al.*, "More diverse means better: Multimodal deep learning meets remote-sensing imagery classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 5, pp. 4340–4354, May 2021.

[108] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, 2006.

[109] Q. Zou, L. Ni, T. Zhang, and Q. Wang, "Deep learning based feature selection for remote sensing scene classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 11, pp. 2321–2325, Nov. 2015.

[110] F. Hu, G.-S. Xia, J. Hu, and L. Zhang, "Transferring deep convolutional neural networks for the scene classification of high-resolution remote sensing imagery," *Remote Sens.*, vol. 7, no. 11, pp. 14680–14707, 2015.

[111] S. Chen and Y. Wang, "Convolutional neural network and convex optimization," Dept. Elect. and Comput. Eng., Univ. of California at San Diego, San Diego, CA, USA, Tech. Rep, 2014. [Online]. Available: https://acsweb.ucsd.edu/ yuw176/report/ECE273.pdf

[112] R. E. Bellman, *Adaptive Control Processes: A Guided Tour*. Princeton, NJ, USA: Princeton Univ. Press, 2015.

[113] G. Hughes, "On the mean accuracy of statistical pattern recognizers," *IEEE Trans. Inf. Theory*, vol. 14, no. 1, pp. 55–63, Jan. 1968.

[114] M. Reichstein *et al.*, "Deep learning and process understanding for data-driven earth system science," *Nature*, vol. 566, no. 7743, pp. 195–204, 2019.

[115] J. M. Bioucas-Dias, A. Plaza, G. Camps-Valls, P. Scheunders, N. Nasrabadi, and J. Chanussot, "Hyperspectral remote sensing data analysis and future challenges," *IEEE Geosci. Remote Sens. Mag.*, vol. 1, no. 2, pp. 6–36, Jun. 2013.

[116] Q. Nguyen and M. Hein, "Optimization landscape and expressivity of deep CNNs," 2017, *arXiv:1710.10928*.

[117] M. Ahmad, S. Shabbir, R. A. Raza, M. Mazzara, S. Distefano, and A. M. Khan, "Artifacts of different dimension reduction methods on hybrid CNN feature hierarchy for hyperspectral image classification," *Optik*, vol. 246, 2021, Art. no. 167757.

[118] M. Ahmad, M. Mazzara, and S. Distefano, "Regularized CNN feature hierarchy for hyperspectral image classification," *Remote Sens.*, vol. 13, no. 12, p. 2275, 2021.

[119] L. Bottou, "Stochastic gradient learning in neural networks," in *Proc. Neuro-Nımes*, vol. 91, no. 8, pp. 1–12, 1991.

[120] N. Qian, "On the momentum term in gradient descent learning algorithms," *Neural Netw.*, vol. 12, no. 1, pp. 145–151, 1999.

[121] G. Hinton, N. Srivastava, and K. Swersky, "Neural networks for machine learning lecture 6a overview of mini-batch gradient descent," vol. 14, no. 8, 2012. [Online]. Available: http://www.cs.toronto.edu/ hinton/coursera/lecture6/lec6.pdf

[122] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. 3rd Int. Conf. Learn. Representations*, San Diego, CA, USA, 2015, *arXiv:1412.6980*.

[123] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," in *Proc. 7th Int. Conf. Learn. Representations*, New Orleans, LA, USA, OpenReview.net, 2019. [Online]. Available: https://openreview.net/forum?id=Bkg6RiCqY7

[124] S. R. Dubey, S. Chakraborty, S. K. Roy, S. Mukherjee, S. K. Singh, and B. B. Chaudhuri, "diffGrad: An optimization method for convolutional neural networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 11, pp. 4500–4511, Nov. 2020.

[125] L. Liu *et al.*, "On the variance of the adaptive learning rate and beyond," in *Proc. Int. Conf. Learn. Representations*, 2019, *arXiv:1908.03265*.

[126] H. Yong, J. Huang, X. Hua, and L. Zhang, "Gradient centralization: A new optimization technique for deep neural networks," in *Proc. Eur. Conf. Comput. Vis.*, Springer, 2020, pp. 635–652.

[127] S. Roy *et al.*, "AngularGrad: A new optimization technique for angular convergence of convolutional neural networks," 2021, *arXiv:2105.10190*.

[128] D. Erhan, Y. Bengio, A. Courville, P.-A. Manzagol, P. Vincent, and S. Bengio, "Why does unsupervised pre-training help deep learning?" *J. Mach. Learn. Res.*, vol. 11, pp. 625–660, 2010.

[129] M. Z. Alom *et al.*, "A state-of-the-art survey on deep learning theory and architectures," *Electronics*, vol. 8, no. 3, p. 292, 2019.

[130] A. Plaza, D. Valencia, and J. Plaza, "An experimental comparison of parallel algorithms for hyperspectral analysis using heterogeneous and homogeneous networks of workstations," *Parallel Comput.*, vol. 34, no. 2, pp. 92–114, 2008.

[131] A. Plaza, J. Plaza, A. Paz, and S. Sanchez, "Parallel hyperspectral image and signal processing [Applications corner]," *IEEE Signal Process. Mag.*, vol. 28, no. 3, pp. 119–126, May 2011.

[132] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.

[133] Y. Bengio, P. Simard, and P. Frasconi, "Learning long-term dependencies with gradient descent is difficult," *IEEE Trans. Neural Netw.*, vol. 5, no. 2, pp. 157–166, Mar. 1994.

[134] Y. Fang *et al.*, "Dimensionality reduction of hyperspectral images based on robust spatial information using locally linear embedding," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 10, pp. 1712–1716, Oct. 2014.

[135] M. Sugiyama, "Dimensionality reduction of multimodal labeled data by local Fisher discriminant analysis," *J. Mach. Learn. Res.*, vol. 8, pp. 1027–1061, 2007.

[136] H.-T. Chen, H.-W. Chang, and T.-L. Liu, "Local discriminant embedding and its variants," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2005, vol. 2, pp. 846–853.

[137] B.-C. Kuo and D. A. Landgrebe, "Nonparametric weighted feature extraction for classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 5, pp. 1096–1105, May 2004.

[138] B. Kumar, O. Dikshit, A. Gupta, and M. K. Singh, "Feature extraction for hyperspectral image classification: A review," *Int. J. Remote Sens.*, vol. 41, no. 16, pp. 6248–6287, 2020.

[139] 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.

[140] Y. Gu, T. Liu, X. Jia, J. A. Benediktsson, and J. Chanussot, "Nonlinear multiple kernel learning with multiple-structure-element extended morphological profiles for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 6, pp. 3235–3247, Jun. 2016.

[141] D. Hong, X. Wu, P. Ghamisi, J. Chanussot, N. Yokoya, and X. X. Zhu, "Invariant attribute profiles: A spatial-frequency joint feature extractor for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 6, pp. 3791–3808, Jun. 2020.

[142] X. Zhang, Y. Sun, K. Jiang, C. Li, L. Jiao, and H. Zhou, "Spatial sequential recurrent neural network for hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 11, pp. 4141–4155, Nov. 2018.

[143] M. Pesaresi and J. A. Benediktsson, "A new approach for the morphological segmentation of high-resolution satellite imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 2, pp. 309–320, Feb. 2001.

[144] Y. Chen, X. Zhao, and X. Jia, "Spectral–spatial classification of hyperspectral data based on deep belief network," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 2381–2392, Jun. 2015.

[145] M. E. Paoletti, J. M. Haut, J. Plaza, and A. Plaza, "Deep&dense convolutional neural network for hyperspectral image classification," *Remote Sens.*, vol. 10, no. 9, p. 1454, 2018.

[146] X. Jin, L. Jie, S. Wang, H. J. Qi, and S. W. Li, "Classifying wheat hyperspectral pixels of healthy heads and fusarium head blight disease using a deep neural network in the wild field," *Remote Sens.*, vol. 10, no. 3, p. 395, 2018.

[147] N. Wu, C. Zhang, X. Bai, X. Du, and Y. He, "Discrimination of chrysanthemum varieties using hyperspectral imaging combined with a deep convolutional neural network," *Molecules*, vol. 23, no. 11, p. 2831, 2018.

[148] Y. Li, W. Xie, and H. Li, "Hyperspectral image reconstruction by deep convolutional neural network for classification," *Pattern Recognit.*, vol. 63, pp. 371–383, 2017.

[149] Y. Zhan, D. Hu, H. Xing, and X. Yu, "Hyperspectral band selection based on deep convolutional neural network and distance density," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 12, pp. 2365–2369, Dec. 2017.

[150] M. E. Paoletti, J. M. Haut, R. Fernandez-Beltran, J. Plaza, A. J. Plaza, and F. Pla, "Deep pyramidal residual networks for spectral–spatial hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 2, pp. 740–754, Feb. 2019.

[151] J. Acquarelli, E. Marchiori, L. Buydens, T. Tran, and T. Van Laarhoven, "Spectral-spatial classification of hyperspectral images: Three tricks and a new learning setting," *Remote Sens.*, vol. 10, no. 7, p. 1156, 2018.

[152] Q. Liu, F. Zhou, R. Hang, and X. Yuan, "Bidirectional-convolutional LSTM based spectral-spatial feature learning for hyperspectral image classification," *Remote Sens.*, vol. 9, no. 12, p. 1330, 2017.

[153] S. K. Roy, S. Manna, T. Song, and L. Bruzzone, "Attention-based adaptive spectral-spatial kernel ResNet for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 9, pp. 7831–7843, Sep. 2021.

[154] B. Liu, X. Yu, P. Zhang, X. Tan, A. Yu, and Z. Xue, "A semi-supervised convolutional neural network for hyperspectral image classification," *Remote Sens. Lett.*, vol. 8, no. 9, pp. 839–848, 2017.

[155] X. Kang, B. Zhuo, and P. Duan, "Semi-supervised deep learning for hyperspectral image classification," *Remote Sens. Lett.*, vol. 10, no. 4, pp. 353–362, 2019.

[156] Y. Wu, G. Mu, C. Qin, Q. Miao, W. Ma, and X. Zhang, "Semi-supervised hyperspectral image classification via spatial-regulated self-training," *Remote Sens.*, vol. 12, no. 1, p. 159, 2020.

[157] Z. Zhang, "Semi-supervised hyperspectral image classification algorithm based on graph embedding and discriminative spatial information," *Microprocessors Microsystems*, 2020, Art. no. 103070.

[158] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.

[159] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.

[160] S. K. Roy, S. Manna, S. R. Dubey, and B. B. Chaudhuri, "LiSHT: Non-parametric linearly scaled hyperbolic tangent activation function for neural networks," 2019, *arXiv:1901.05894*.

[161] T. Williams and R. Li, "Wavelet pooling for convolutional neural networks," in *Proc. 24th Int. Conf. Artificial Intell. Statistics*, PMLR, vol. 130, Apr. 2021, pp. 1936–1944.

[162] J. T. Springenberg, A. Dosovitskiy, T. Brox, and M. Riedmiller, "Striving for simplicity: The all convolutional net," 2014, *arXiv:1412.6806*.

[163] D. H. Hubel and T. N. Wiesel, "Receptive fields, binocular interaction and functional architecture in the cat's visual cortex," *J. Physiol.*, vol. 160, no. 1, pp. 106–154, 1962.

[164] K. Fukushima, "Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position," *Biol. Cybern.*, vol. 36, no. 4, pp. 193–202, 1980.

[165] A. Voulodimos, N. Doulamis, A. Doulamis, and E. Protopapadakis, "Deep learning for computer vision: A brief review," *Comput. Intell. Neurosci.*, vol. 2018, 2018, Art. no. 7068349.

[166] J. Gu *et al.*, "Recent advances in convolutional neural networks," *Pattern Recognit.*, vol. 77, pp. 354–377, 2018.

[167] M. Lin, Q. Chen, and S. Yan, "Network in network," 2013, *arXiv:1312.4400*.

[168] H. Gao, Y. Yang, C. Li, H. Zhou, and X. Qu, "Joint alternate small convolution and feature reuse for hyperspectral image classification," *ISPRS Int. J. Geo- Inf.*, vol. 7, no. 9, p. 349, 2018.

[169] W. Zhao *et al.*, "Superpixel-based multiple local CNN for panchromatic and multispectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 7, pp. 4141–4156, Jul. 2017.

[170] H. Alhichri, N. Alajlan, Y. Bazi, and T. Rabczuk, "Multi-scale convolutional neural network for remote sensing scene classification," in *Proc. IEEE Int. Conf. Electro/Inf. Technol.*, 2018, pp. 1–5.

[171] M. Noor, S. Salwa, J. Ren, S. Marshall, and K. Michael, "Hyperspectral image enhancement and mixture deep-learning classification of corneal epithelium injuries," *Sensors*, vol. 17, no. 11, p. 2644, 2017.

[172] J. Leng, T. Li, G. Bai, Q. Dong, and H. Dong, "Cube-CNN-SVM: A novel hyperspectral image classification method," in *Proc. IEEE 28th Int. Conf. Tools Artif. Intell.*, 2016, pp. 1027–1034.

[173] S. Yu, S. Jia, and C. Xu, "Convolutional neural networks for hyperspectral image classification," *Neurocomputing*, vol. 219, pp. 88–98, 2017.

[174] H. Wu and S. Prasad, "Convolutional recurrent neural networks for hyperspectral data classification," *Remote Sens.*, vol. 9, no. 3, p. 298, 2017.

[175] Z. Qiu, J. Chen, Y. Zhao, S. Zhu, Y. He, and C. Zhang, "Variety identification of single rice seed using hyperspectral imaging combined with convolutional neural network," *Appl. Sci.*, vol. 8, no. 2, p. 212, 2018.

[176] Q. Huang, W. Li, and X. Xie, "Convolutional neural network for medical hyperspectral image classification with kernel fusion," in *Proc. Int. Conf. Biol. Inf. Biomed. Eng.*, VDE, 2018, pp. 1–4.

[177] K. Charmisha, V. Sowmya, and K. Soman, "Dimensionally reduced features for hyperspectral image classification using deep learning," in *Proc. Int. Conf. Commun. Cyber Phys. Eng.*, Springer, 2018, pp. 171–179.

[178] G. Turra, S. Arrigoni, and A. Signoroni, "CNN-based identification of hyperspectral bacterial signatures for digital microbiology," in *Proc. Int. Conf. Image Anal. Process.*, Springer, 2017, pp. 500–510.

[179] J. Li, X. Zhao, Y. Li, Q. Du, B. Xi, and J. Hu, "Classification of hyperspectral imagery using a new fully convolutional neural network," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 2, pp. 292–296, Feb. 2018.

[180] J. M. Haut, M. E. Paoletti, J. Plaza, A. Plaza, and J. Li, "Hyperspectral image classification using random occlusion data augmentation," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 11, pp. 1751–1755, Nov. 2019.

[181] Y. Xu, B. Du, F. Zhang, and L. Zhang, "Hyperspectral image classification via a random patches network," *ISPRS J. Photogrammetry Remote Sens.*, vol. 142, pp. 344–357, 2018.

[182] Y. Wang, T. Song, Y. Xie, and S. K. Roy, "A probabilistic neighbourhood pooling-based attention network for hyperspectral image classification," *Remote Sens. Lett.*, vol. 13, no. 1, pp. 65–75, 2021.

[183] C. Ding, Y. Li, Y. Xia, W. Wei, L. Zhang, and Y. Zhang, "Convolutional neural networks based hyperspectral image classification method with adaptive kernels," *Remote Sens.*, vol. 9, no. 6, p. 618, 2017.

[184] Y. Chen, L. Zhu, P. Ghamisi, X. Jia, G. Li, and L. Tang, "Hyperspectral images classification with Gabor filtering and convolutional neural network," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 12, pp. 2355–2359, Dec. 2017.

[185] J. Zhu, L. Fang, and P. Ghamisi, "Deformable convolutional neural networks for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 8, pp. 1254–1258, Aug. 2018.

[186] L. Ran, Y. Zhang, W. Wei, and Q. Zhang, "A hyperspectral image classification framework with spatial pixel pair features," *Sensors*, vol. 17, no. 10, p. 2421, 2017.

[187] Z. Zhong, J. Li, Z. Luo, and M. Chapman, "Spectral-spatial residual network for hyperspectral image classification: A 3-D deep learning framework," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 2, pp. 847–858, Feb. 2018.

[188] M. Paoletti, J. Haut, J. Plaza, and A. Plaza, "A new deep convolutional neural network for fast hyperspectral image classification," *ISPRS J. Photogrammetry Remote Sens.*, vol. 145, pp. 120–147, 2018.

[189] S. Li, X. Zhu, Y. Liu, and J. Bao, "Adaptive spatial–spectral feature learning for hyperspectral image classification," *IEEE Access*, vol. 7, pp. 61534–61547, 2019.

[190] S. K. Roy, M. E. Paoletti, J. M. Haut, E. M. T. Hendrix, and A. Plaza, "A new max–min convolutional network for hyperspectral image classification," in *Proc. 11th Workshop Hyperspectral Imag. Signal Process.: Evol. Remote Sens.*, 2021, pp. 1–5.

[191] M. E. Paoletti, J. M. Haut, S. K. Roy, and E. M. Hendrix, "Rotation equivariant convolutional neural networks for hyperspectral image classification," *IEEE Access*, vol. 8, pp. 179575–179591, 2020.

[192] H. Zhang, Y. Li, Y. Jiang, P. Wang, Q. Shen, and C. Shen, "Hyperspectral classification based on lightweight 3-D-CNN with transfer learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 8, pp. 5813–5828, Aug. 2019.

[193] S. Jia *et al.*, "A lightweight convolutional neural network for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 5, pp. 4150–4163, May 2021.

[194] S. K. Roy, S. Chatterjee, S. Bhattacharyya, B. B. Chaudhuri, and J. Platoš, "Lightweight spectral–spatial squeeze-and-excitation residual bag-of-features learning for hyperspectral classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 8, pp. 5277–5290, Aug. 2020.

[195] S. K. Roy, R. Mondal, M. E. Paoletti, J. M. Haut, and A. Plaza, "Morphological convolutional neural networks for hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 8689–8702, 2021.

[196] Y. Li, H. Zhang, and Q. Shen, "Spectral–spatial classification of hyperspectral imagery with 3D convolutional neural network," *Remote Sens.*, vol. 9, no. 1, p. 67, 2017.

[197] 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.

[198] L. Jiao, M. Liang, H. Chen, S. Yang, H. Liu, and X. Cao, "Deep fully convolutional network-based spatial distribution prediction for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 10, pp. 5585–5599, Oct. 2017.

[199] 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.

[200] M. He, B. Li, and H. Chen, "Multi-scale 3D deep convolutional neural network for hyperspectral image classification," in *Proc. IEEE Int. Conf. Image Process.*, 2017, pp. 3904–3908.

[201] H. Dong, L. Zhang, and B. Zou, "Band attention convolutional networks for hyperspectral image classification," 2019, *arXiv:1906.04379*.

[202] N. He *et al.*, "Feature extraction with multiscale covariance maps for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 2, pp. 755–769, Feb. 2019.

[203] G. Cheng, Z. Li, J. Han, X. Yao, and L. Guo, "Exploring hierarchical convolutional features for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 11, pp. 6712–6722, Nov. 2018.

[204] Z. Gong, P. Zhong, Y. Yu, W. Hu, and S. Li, "A CNN with multiscale convolution and diversified metric for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 6, pp. 3599–3618, Jun. 2019.

[205] P. Zhong, N. Peng, and R. Wang, "Learning to diversify patch-based priors for remote sensing image restoration," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 11, pp. 5225–5245, Nov. 2015.

[206] L. Liu, Z. Shi, B. Pan, N. Zhang, H. Luo, and X. Lan, "Multiscale deep spatial feature extraction using virtual RGB image for hyperspectral imagery classification," *Remote Sens.*, vol. 12, no. 2, p. 280, 2020.

[207] X. Ma, A. Fu, J. Wang, H. Wang, and B. Yin, "Hyperspectral image classification based on deep deconvolution network with skip architecture," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 8, pp. 4781–4791, Aug. 2018.

[208] A. Sellami, M. Farah, I. R. Farah, and B. Solaiman, "Hyperspectral imagery classification based on semi-supervised 3-D deep neural network and adaptive band selection," *Expert Syst. Appl.*, vol. 129, pp. 246–259, 2019.

[209] S. K. Roy, S. Das, T. Song, and B. Chanda, "DARecNet-BS: Unsupervised dual-attention reconstruction network for hyperspectral band selection," *IEEE Geosci. Remote Sens. Lett.*, vol. 18, no. 12, pp. 2152–2156, Dec. 2021.

[210] S. Mei, J. Ji, Y. Geng, Z. Zhang, X. Li, and Q. Du, "Unsupervised spatial-spectral feature learning by 3D convolutional autoencoder for hyperspectral classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 9, pp. 6808–6820, Sep. 2019.

[211] S. K. Roy, D. Hong, P. Kar, X. Wu, X. Liu, and D. Zhao, "Lightweight heterogeneous kernel convolution for hyperspectral image classification with noisy labels," *IEEE Geosci. Remote Sens. Lett.*, to be published, doi: 10.1109/LGRS.2021.3112755.

[212] 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.

[213] 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.

[214] M. E. Paoletti, J. M. Haut, T. Alipour-Fard, S. K. Roy, E. M. Hendrix, and A. Plaza, "Separable attention network in single-and mixed-precision floating point for land-cover classification of remote sensing images," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022, Art no. 2501605, doi: 10.1109/LGRS.2021.3108965.

[215] S. K. Roy, P. Kar, D. Hong, X. Wu, A. Plaza, and J. Chanussot, "Revisiting deep hyperspectral feature extraction networks via gradient centralized convolution," *IEEE Trans. Geosci. Remote Sens.*, to be published, doi: 10.1109/TGRS.2021.3120198.

[216] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," in *Proc. Int. Conf. Learn. Representations*, 2016, *arXiv:1609.02907*.

[217] D. Hong, L. Gao, J. Yao, B. Zhang, A. Plaza, and J. Chanussot, "Graph convolutional networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 7, pp. 5966–5978, Jul. 2021.

[218] A. Qin, Z. Shang, J. Tian, Y. Wang, T. Zhang, and Y. Y. Tang, "Spectral–spatial graph convolutional networks for semisupervised hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 2, pp. 241–245, Feb. 2019.

[219] S. Wan, C. Gong, P. Zhong, S. Pan, G. Li, and J. Yang, "Hyperspectral image classification with context-aware dynamic graph convolutional network," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 1, pp. 597–612, Jan. 2021.

[220] J. Zhu, L. Wu, H. Hao, X. Song, and Y. Lu, "Auto-encoder based for high spectral dimensional data classification and visualization," in *Proc. IEEE 2nd Int. Conf. Data Sci. Cyberspace*, 2017, pp. 350–354.

[221] A. Hassanzadeh, A. Kaarna, and T. Kauranne, "Unsupervised multi-manifold classification of hyperspectral remote sensing images with contractive autoencoder," in *Proc. Scand. Conf. Image Anal.*, Springer, 2017, pp. 169–180.

[222] Y. Wang, Y. Jiang, Y. Wu, and Z.-H. Zhou, "Multi-manifold clustering," in *Proc. Pacific Rim Int. Conf. Artif. Intell.*, Springer, 2010, pp. 280–291.

[223] S. Rifai, P. Vincent, X. Muller, X. Glorot, and Y. Bengio, "Contractive auto-encoders: Explicit invariance during feature extraction," in *Proc. 28th Int. Conf. Int. Conf. Mach. Learn.*, Madison, WI, USA, Omnipress, 2011, pp. 833–840.

[224] X. Zhang, Y. Liang, C. Li, N. Huyan, L. Jiao, and H. Zhou, "Recursive autoencoders-based unsupervised feature learning for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 11, pp. 1928–1932, Nov. 2017.

[225] 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.

[226] K. He, J. Sun, and X. Tang, "Guided image filtering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 6, pp. 1397–1409, Jun. 2013.

[227] X. Sun, F. Zhou, J. Dong, F. Gao, Q. Mu, and X. Wang, "Encoding spectral and spatial context information for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 12, pp. 2250–2254, Dec. 2017.

[228] C. Zhao, X. Wan, G. Zhao, B. Cui, W. Liu, and B. Qi, "Spectral–spatial classification of hyperspectral imagery based on stacked sparse autoencoder and random forest," *Eur. J. remote Sens.*, vol. 50, no. 1, pp. 47–63, 2017.

[229] X. Wan, C. Zhao, Y. Wang, and W. Liu, "Stacked sparse autoencoder in hyperspectral data classification using spectral-spatial, higher order statistics and multifractal spectrum features," *Infrared Phys. Technol.*, vol. 86, pp. 77–89, 2017.

[230] F. Lv, M. Han, and T. Qiu, "Remote sensing image classification based on ensemble extreme learning machine with stacked autoencoder," *IEEE Access*, vol. 5, pp. 9021–9031, 2017.

[231] M. Ahmad, A. M. Khan, M. Mazzara, and S. Distefano, "Multi-layer extreme learning machine-based autoencoder for hyperspectral image classification," in *Proc. 14th Int. Conf. Comput. Vis. Theory Appl.*, Prague, Czech Republic, 2019, pp. 25–27.

[232] P. Zhou, J. Han, G. Cheng, and B. Zhang, "Learning compact and discriminative stacked autoencoder for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 7, pp. 4823–4833, Jul. 2019.

[233] R. Lan, Z. Li, Z. Liu, T. Gu, and X. Luo, "Hyperspectral image classification using k-sparse denoising autoencoder and spectral-restricted spatial characteristics," *Appl. Soft Comput.*, vol. 74, pp. 693–708, 2019.

[234] S. Paul and D. N. Kumar, "Spectral–spatial classification of hyperspectral data with mutual information based segmented stacked autoencoder approach," *ISPRS J. Photogrammetry Remote Sens.*, vol. 138, pp. 265–280, 2018.

[235] B. Liu, Q. Zhang, L. Ying, W. Chang, and M. Zhou, "Spatial–spectral jointed stacked auto-encoder-based deep learning for oil slick extraction from hyperspectral images," *J. Indian Soc. Remote Sens.*, vol. 47, no. 12, pp. 1989–1997, 2019.

[236] G. E. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," *Neural Comput.*, vol. 18, no. 7, pp. 1527–1554, 2006.

[237] N. Zhang, S. Ding, J. Zhang, and Y. Xue, "An overview on restricted Boltzmann machines," *Neurocomputing*, vol. 275, pp. 1186–1199, 2018.

[238] B. Ayhan and C. Kwan, "Application of deep belief network to land cover classification using hyperspectral images," in *Proc. Adv. Neural Netw.*, F. Cong, A. Leung, and Q. Wei, Eds., Cham, Switzerland, Springer, 2017, pp. 269–276.

[239] U. Shaham *et al.*, "A deep learning approach to unsupervised ensemble learning," in *Proc. Int. Conf. Mach. Learn.*, 2016, pp. 30–39.

[240] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov, "Improving neural networks by preventing co-adaptation of feature detectors," 2012, *arXiv:1207.0580*.

[241] H. Xiong, A. J. Rodríguez-Sánchez, S. Szedmak, and J. Piater, "Diversity priors for learning early visual features," *Frontier Comput. Neurosci.*, vol. 9, p. 104, 2015, doi: 10.3389/fncom.2015.00104.

[242] P. Zhong, Z. Gong, S. Li, and C.-B. Schönlieb, "Learning to diversify deep belief networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 6, pp. 3516–3530, Jun. 2017.

[243] J. Li, B. Xi, Y. Li, Q. Du, and K. Wang, "Hyperspectral classification based on texture feature enhancement and deep belief networks," *Remote Sens.*, vol. 10, no. 3, p. 396, 2018.

[244] K. Tan, F. Wu, Q. Du, P. Du, and Y. Chen, "A parallel Gaussian–Bernoulli restricted Boltzmann machine for mining area classification with hyperspectral imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 2, pp. 627–636, Feb. 2019.

[245] S. Li, W. Song, L. Fang, Y. Chen, P. Ghamisi, and J. A. Benediktsson, "Deep learning for hyperspectral image classification: An overview," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 9, pp. 6690–6709, Sep. 2019.

[246] A. Sellami and I. Farah, "Spectra–spatial graph-based deep restricted Boltzmann networks for hyperspectral image classification," in *Proc. Photon. Electromagn. Res. Symp.*, 2019, pp. 1055–1062.

[247] 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.

[248] M. E. Paoletti, J. M. Haut, J. Plaza, and A. Plaza, "Scalable recurrent neural network for hyperspectral image classification," *J. Supercomput.*, vol. 76, pp. 8866–8882, 2020.

[249] F. Zhou, R. Hang, Q. Liu, and X. Yuan, "Hyperspectral image classification using spectral–spatial LSTMs," *Neurocomputing*, vol. 328, pp. 39–47, 2019.

[250] A. Sharma, X. Liu, and X. Yang, "Land cover classification from multi-temporal, multi-spectral remotely sensed imagery using patch-based recurrent neural networks," *Neural Netw.*, vol. 105, pp. 346–355, 2018.

[251] H. Wu and S. Prasad, "Semi-supervised deep learning using pseudo labels for hyperspectral image classification," *IEEE Trans. Image Process.*, vol. 27, no. 3, pp. 1259–1270, Mar. 2018.

[252] F. Zhou, R. Hang, Q. Liu, and X. Yuan, "Integrating convolutional neural network and gated recurrent unit for hyperspectral image spectral–spatial classification," in *Proc. Chin. Conf. Pattern Recognit. Comput. Vis.*, Springer, 2018, pp. 409–420.

[253] H. Luo, "Shorten spatial-spectral RNN with parallel-GRU for hyperspectral image classification," 2018, *arXiv:1810.12563*.

[254] C. Shi and C.-M. Pun, "Multi-scale hierarchical recurrent neural networks for hyperspectral image classification," *Neurocomputing*, vol. 294, pp. 82–93, 2018.

[255] X. Yang, Y. Ye, X. Li, R. Y. Lau, X. Zhang, and X. Huang, "Hyperspectral image classification with deep learning models," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 9, pp. 5408–5423, Sep. 2018.

[256] M. Seydgar, A. A. Naeini, M. Zhang, W. Li, and M. Satari, "3-D convolution-recurrent networks for spectral–spatial classification of hyperspectral images," *Remote Sens.*, vol. 11, no. 7, p. 883, 2019.

[257] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *J. Big Data*, vol. 6, no. 1, pp. 1–48, 2019.

[258] S. Jia, S. Jiang, Z. Lin, N. Li, M. Xu, and S. Yu, "A survey: Deep learning for hyperspectral image classification with few labeled samples," *Neurocomputing*, vol. 448, pp. 179–204, 2021.

[259] X. Yu, X. Wu, C. Luo, and P. Ren, "Deep learning in remote sensing scene classification: A data augmentation enhanced convolutional neural network framework," *GIScience Remote Sens.*, vol. 54, no. 5, pp. 741–758, 2017.

[260] W. Li, C. Chen, M. Zhang, H. Li, and Q. Du, "Data augmentation for hyperspectral image classification with deep CNN," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 4, pp. 593–597, Apr. 2019.

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

[262] J. F. R. Rochac, N. Zhang, L. Thompson, and T. Oladunni, "A data augmentation-assisted deep learning model for high dimensional and highly imbalanced hyperspectral imaging data," in *Proc. 9th Int. Conf. Inf. Sci. Technol.*, 2019, pp. 362–367.

[263] J. Nalepa, M. Myller, and M. Kawulok, "Training-and test-time data augmentation for hyperspectral image segmentation," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 2, pp. 292–296, Feb. 2020.

[264] J. Nalepa, M. Myller, and M. Kawulok, "Hyperspectral data augmentation," 2019, *arXiv:1903.05580*.

[265] J. E. Van Engelen and H. H. Hoos, "A survey on semi-supervised learning," *Mach. Learn.*, vol. 109, no. 2, pp. 373–440, 2020.

[266] N. N. Pise and P. Kulkarni, "A survey of semi-supervised learning methods," in *Proc. Int. Conf. Comput. Intell. Secur.*, 2008, vol. 2, pp. 30–34.

[267] S. S. Sawant and M. Prabukumar, "Semi-supervised techniques based hyper-spectral image classification: A survey," in *Proc. Innov. Power Adv. Comput. Technol.*, 2017, pp. 1–8.

[268] B. Fang, Y. Li, H. Zhang, and J. C.-W. Chan, "Semi-supervised deep learning classification for hyperspectral image based on dual-strategy sample selection," *Remote Sens.*, vol. 10, no. 4, p. 574, 2018.

[269] S. Zhou, Z. Xue, and P. Du, "Semisupervised stacked autoencoder with cotraining for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 6, pp. 3813–3826, Jun. 2019.

[270] F. Li, D. A. Clausi, L. Xu, and A. Wong, "ST-IRGS: A region-based self-training algorithm applied to hyperspectral image classification and segmentation," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 1, pp. 3–16, Jan. 2018.

[271] M. S. Aydemir and G. Bilgin, "Semisupervised hyperspectral image classification using small sample sizes," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 5, pp. 621–625, May 2017.

[272] I. Goodfellow *et al.*, "Generative adversarial nets," in *Proc. Adv. neural Inf. Process. Syst.*, 2014, pp. 2672–2680.

[273] Y. Zhan, D. Hu, Y. Wang, and X. Yu, "Semisupervised hyperspectral image classification based on generative adversarial networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 2, pp. 212–216, Feb. 2018.

[274] Z. He, H. Liu, Y. Wang, and J. Hu, "Generative adversarial networks-based semi-supervised learning for hyperspectral image classification," *Remote Sens.*, vol. 9, no. 10, p. 1042, 2017.

[275] 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.

[276] Y. Zhan *et al.*, "Semi-supervised classification of hyperspectral data based on generative adversarial networks and neighborhood majority voting," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2018, pp. 5756–5759.

[277] J. Feng, H. Yu, L. Wang, X. Cao, X. Zhang, and L. Jiao, "Classification of hyperspectral images based on multiclass spatial-spectral generative adversarial networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 8, pp. 5329–5343, Aug. 2019.

[278] Z. Zhong, J. Li, D. A. Clausi, and A. Wong, "Generative adversarial networks and conditional random fields for hyperspectral image classification," *IEEE Trans. Cybern.*, vol. 50, no. 7, pp. 3318–3329, Jul. 2020.

[279] X. Wang, K. Tan, Q. Du, Y. Chen, and P. Du, "Caps-TripleGan: GAN-assisted CapsNet for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 9, pp. 7232–7245, Sep. 2019.

[280] Z. Xue, "Semi-supervised convolutional generative adversarial network for hyperspectral image classification," *IET Image Process.*, vol. 14, no. 4, pp. 709–719, 2019.

[281] W.-Y. Wang, H.-C. Li, Y.-J. Deng, L.-Y. Shao, X.-Q. Lu, and Q. Du, "Generative adversarial capsule network with ConvLSTM for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 18, no. 3, pp. 523–527, Mar. 2021.

[282] T. Alipour-Fard and H. Arefi, "Structure aware generative adversarial networks for hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 5424–5438, 2020.

[283] S. K. Roy, J. M. Haut, M. E. Paoletti, S. R. Dubey, and A. Plaza, "Generative adversarial minority oversampling for spectral-spatial hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–15, 2022.

[284] J. Yang, Y.-Q. Zhao, and J. C.-W. Chan, "Learning and transferring deep joint spectral-spatial features for hyperspectral classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 8, pp. 4729–4742, Aug. 2017.

[285] L. Windrim, A. Melkumyan, R. J. Murphy, A. Chlingaryan, and R. Ramakrishnan, "Pretraining for hyperspectral convolutional neural network classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 5, pp. 2798–2810, May 2018.

[286] X. Liu, Q. Sun, Y. Meng, M. Fu, and S. Bourennane, "Hyperspectral image classification based on parameter-optimized 3D-CNNs combined with transfer learning and virtual samples," *Remote Sens.*, vol. 10, no. 9, p. 1425, 2018.

[287] O. Day and T. M. Khoshgoftaar, "A survey on heterogeneous transfer learning," *J. Big Data*, vol. 4, no. 1, pp. 1–42, 2017.

[288] J. Lin, R. Ward, and Z. J. Wang, "Deep transfer learning for hyperspectral image classification," in *Proc. IEEE 20th Int. Workshop Multimedia Signal Process.*, 2018, pp. 1–5.

[289] X. Li, L. Zhang, B. Du, L. Zhang, and Q. Shi, "Iterative reweighting heterogeneous transfer learning framework for supervised remote sensing image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 5, pp. 2022–2035, May 2017.

[290] Y. Liu and C. Xiao, "Transfer learning for hyperspectral image classification using convolutional neural network," in *Proc. SPIE*, vol. 11432, 2020, Art. no. 114320E.

[291] J. Lin, C. He, Z. J. Wang, and S. Li, "Structure preserving transfer learning for unsupervised hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 10, pp. 1656–1660, Oct. 2017.

[292] R. P. de Lima and K. Marfurt, "Convolutional neural network for remote-sensing scene classification: Transfer learning analysis," *Remote Sens.*, vol. 12, no. 1, p. 86, 2020.

[293] R. Ganti and A. Gray, " UPAL: Unbiased pool based active learning," in *Proc. Artif. Intell. Statist.*, 2012, pp. 422–431.

[294] P. Melville and R. J. Mooney, "Diverse ensembles for active learning," in *Proc. 21st Int. Conf. Mach. Learn.*, 2004, pp. 584–591.

[295] C. C. Aggarwal, X. Kong, Q. Gu, J. Han, and S. Y. Philip, "Active learning: A survey," in *Data Classification: Algorithms and Applications*. Boca Raton, FL, USA: CRC Press, 2014, pp. 571–605.

[296] H. S. Seung, M. Opper, and H. Sompolinsky, "Query by committee," in *Proc. 5th Annu. Workshop Comput. Learn. Theory*, 1992, pp. 287–294.

[297] B. Settles, "Active learning literature survey," Univ. of Wisconsin-Madison Dept. of Comput. Sci., Tech. Rep. 1648, 2009.

[298] C. Liu, L. He, Z. Li, and J. Li, "Feature-driven active learning for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 1, pp. 341–354, Jan. 2017.

[299] Y. Zhang, G. Cao, X. Li, B. Wang, and P. Fu, "Active semi-supervised random forest for hyperspectral image classification," *Remote Sens.*, vol. 11, no. 24, p. 2974, 2019.

[300] J. Guo, X. Zhou, J. Li, A. Plaza, and S. Prasad, "Superpixel-based active learning and online feature importance learning for hyperspectral image analysis," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 1, pp. 347–359, Jan. 2017.

[301] Z. Xue, S. Zhou, and P. Zhao, "Active learning improved by neighborhoods and superpixels for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 3, pp. 469–473, Mar. 2018.

[302] K. Bhardwaj, A. Das, and S. Patra, "Spectral–spatial active learning with attribute profile for hyperspectral image classification," in *Proc. Int. Conf. Intell. Comput. Smart Commun.*, Springer, 2020, pp. 1219–1229.

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

[304] Z. Zhang and M. M. Crawford, "A batch-mode regularized multimetric active learning framework for classification of hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 11, pp. 6594–6609, Nov. 2017.

[305] X. Xu, J. Li, and S. Li, "Multiview intensity-based active learning for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 2, pp. 669–680, Feb. 2018.

[306] M. K. Pradhan, S. Minz, and V. K. Shrivastava, "Fisher discriminant ratio based multiview active learning for the classification of remote sensing images," in *Proc. 4th Int. Conf. Recent Adv. Inf. Technol.*, 2018, pp. 1–6.

[307] Z. Zhang, E. Pasolli, and M. M. Crawford, "An adaptive multiview active learning approach for spectral–spatial classification of hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 4, pp. 2557–2570, Apr. 2020.

[308] Y. Li, T. Lu, and S. Li, "Subpixel-pixel-superpixel-based multiview active learning for hyperspectral images classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 7, pp. 4976–4988, Jul. 2020.

[309] Y. Sun, J. Li, W. Wang, A. Plaza, and Z. Chen, "Active learning based autoencoder for hyperspectral imagery classification," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2016, pp. 469–472.

[310] P. Liu, H. Zhang, and K. B. Eom, "Active deep learning for classification of hyperspectral images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 2, pp. 712–724, Feb. 2017.

[311] J. M. Haut, M. E. Paoletti, J. Plaza, J. Li, and A. Plaza, "Active learning with convolutional neural networks for hyperspectral image classification using a new Bayesian approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 11, pp. 6440–6461, Nov. 2018.

[312] J. Lin, L. Zhao, S. Li, R. Ward, and Z. J. Wang, "Active-learning-incorporated deep transfer learning for hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 11, pp. 4048–4062, Nov. 2018.

[313] C. Deng, Y. Xue, X. Liu, C. Li, and D. Tao, "Active transfer learning network: A unified deep joint spectral-spatial feature learning model for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 3, pp. 1741–1754, Mar. 2018.

[314] C. Deng, X. Liu, C. Li, and D. Tao, "Active multi-kernel domain adaptation for hyperspectral image classification," *Pattern Recognit.*, vol. 77, pp. 306–315, 2018.

[315] M. Ahmad, R. A. Raza, and M. Mazzara, "Multiclass non-randomized spectral-spatial active learning for hyperspectral image classification," *Appl. Sci.*, vol. 10, p. 4739, 2020.

[316] R. O. Green *et al.*, "Imaging spectroscopy and the airborne visible/infrared imaging spectrometer (AVIRIS)," *Remote Sens. Environ.*, vol. 65, no. 3, pp. 227–248, 1998.

[317] X. Huang and L. Zhang, "A comparative study of spatial approaches for urban mapping using hyperspectral ROSIS images over Pavia City, Northern Italy," *Int. J. Remote Sens.*, vol. 30, no. 12, pp. 3205–3221, 2009.

[318] X. Xu, J. Li, and A. Plaza, "Fusion of hyperspectral and LiDAR data using morphological component analysis," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2016, pp. 3575–3578.

[319] X. Xu, W. Li, Q. Ran, Q. Du, L. Gao, and B. Zhang, "Multisource remote sensing data classification based on convolutional neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 2, pp. 937–949, Feb. 2018.

[320] J. Li, J. M. Bioucas-Dias, and A. Plaza, "Semisupervised hyperspectral image segmentation using multinomial logistic regression with active learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 11, pp. 4085–4098, Nov. 2010.

[321] 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.
[322] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.
[323] K. Cho, B. Van Merriënboer, D. Bahdanau, and Y. Bengio, "On the properties of neural machine translation: Encoder–decoder approaches," in *Proc. 8th Workshop Syntax, Semantics Structure Statist. Transl.*, 2014, pp. 103–111.
[324] 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.
[325] A. Ben Hamida, A. Benoit, P. Lambert, and C. Ben Amar, "3-D deep learning approach for remote sensing image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 8, pp. 4420–4434, Aug. 2018.
[326] J. Li, J. M. Bioucas-Dias, and A. Plaza, "Hyperspectral image segmentation using a new Bayesian approach with active learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 10, pp. 3947–3960, Oct. 2011.
[327] J. Haut, M. Paoletti, A. Paz-Gallardo, J. Plaza, A. Plaza, and J. Vigo-Aguiar, "Cloud implementation of logistic regression for hyperspectral image classification," in *Proc. 17th Int. Conf. Comput. Math. Methods Sci. Eng.*, Cádiz, Spain:Costa Ballena (Rota), 2017, vol. 3, pp. 1063–2321.
[328] L. Windrim, R. Ramakrishnan, A. Melkumyan, R. J. Murphy, and A. Chlingaryan, "Unsupervised feature-learning for hyperspectral data with autoencoders," *Remote Sens.*, vol. 11, no. 7, p. 864, 2019.
[329] X. Chen, M. Li, and Y. Xiaoquan, "Stacked denoise autoencoder based feature extraction and classification for hyperspectral images," *J. Sensors*, vol. 2016, pp. 1–10, 2016.

Muhammad Ahmad received the M.S. degree in electronics engineering from International Islamic University, Islamabad, Pakistan, in 2011, and the Ph.D. degree in computer science and engineering from Innopolis University, Innopolis, Russia, in 2019, and the second Ph.D. degree in cyber-physical systems from the University of Messina, Messina, Italy, in 2021.

He is currently associated with the National University of Computer and Emerging Sciences (FAST-NUCES), Islamabad, Pakistan. He was an Assistant Professor, Lecturer, Instructor, Research Fellow, Research Associate, and Research Assistant with a number of international/national universities. He was with Ericsson (Mobilink Project) as Radio Access Network (RAN) Supervisor. He has authored and co-authored more than 70 scientific contributions to international journals, conferences, and books. He has supervised/co-supervised several graduates (M.S. and Ph.D.). His research interests include hyperspectral imaging, remote sensing, machine learning, computer vision, and wearable computing.

Dr. Ahmad served/serving as a Lead/Guest Editor for several special issues in journals SCI/E and JCR. He has delivered a number of invited and keynote talks and reviewed (reviewing) the technology-leading articles for journals.

Sidrah Shabir received the Bachelor's degree in computer engineering from COMSATS University Islamabad, Islamabad, Pakistan, in 2015, and the Master's degree in computer engineering from the Khwaja Fareed University of Engineering and Information Technology, Rahim Yar Khan, Pakistan, in 2020.

She is currently a Lab Engineer with the Department of Computer Engineering, Khwaja Fareed University of Engineering and Information Technology. Her research interests include machine learning, hyperspectral imaging, and hardware accelerator design for machine learning.

Swalpa Kumar Roy (Student Member, IEEE) received the Bachelor's degree from the West Bengal University of Technology, Kolkata, India, in 2012, the Master's degree from Indian Institute of Engineering Science and Technology (IIEST Shibpur), Shibpur, Howrah, India, in 2015, and the Ph.D. degree from the University of Calcutta, Kolkata, India, in 2021, all in computer science and engineering.

From July 2015 to March 2016, he was a Project Linked Person with the Optical Character Recognition (OCR) Laboratory, Computer Vision and Pattern Recognition Unit, Indian Statistical Institute, Kolkata, India. He is currently an Assistant Professor with the Department of Computer Science and Engineering, Jalpaiguri Government Engineering College, West Bengal, India. His research interests include computer vision, deep learning, and remote sensing.

Dr. Roy was nominated for the Indian National Academy of Engineering (INAE) engineering teachers mentoring fellowship program by INAE Fellows in 2021 and also a recipient of the Outstanding Paper Award in second Hyperspectral Sensing Meets Machine Learning and Pattern Analysis (HyperMLPA) at the Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing (WHISPERS) in 2021. He has served as a Reviewer for IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING and *IEEE Geoscience and Remote Sensing Letters*.

Danfeng Hong (Senior Member, IEEE) received the M.Sc. degree (Summa Cum Laude) in computer vision from the College of Information Engineering, Qingdao University, Qingdao, China, in 2015, the Dr.-Ing. degree (Summa Cum Laude) from the Signal Processing in Earth Observation (SiPEO), Technical University of Munich (TUM), Munich, Germany, in 2019.

From 2015 to 2019, he was a Research Associate with the Remote Sensing Technology Institute (IMF), German Aerospace Center (DLR), Oberpfaffenhofen, Germany. Since 2019, he has been a Research Scientist and led a Spectral Vision Working Group, IMF, DLR. He has also been an Adjunct Scientist with GIPSA-Lab, Grenoble INP, CNRS, Univ. Grenoble Alpes, Grenoble, France, since 2020. He is currently with the Key Laboratory of Digital Earth Science, Aerospace Information Research Institute (AIR), Chinese Academy of Sciences (CAS), Beijing, China. His research interests include signal/image processing and analysis, hyperspectral remote sensing, machine/deep learning, artificial intelligence, and their applications in Earth Vision.

Dr. Hong is an Editorial Board Member of Remote Sensing and a Topical Associate Editor of IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING. He was a recipient of the Best Reviewer Award of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING in 2021 and the Jose Bioucas Dias Award for the outstanding paper at the Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing (WHISPERS) in 2021. He is also a Leading Guest Editor of *International Journal of Applied Earth Observation and Geoinformation*, *IEEE Journal of Selected Topics in Applied Earth Observations*, and *Remote Sensing*.

Xin Wu (Member, IEEE) received the M.Sc. degree in computer science and technology from the College of Information Engineering, Qingdao University, Qingdao, China, in 2014, and the Ph.D. degree from the School of Information and Electronics, Beijing Institute of Technology (BIT), Beijing, China, in 2020.

In 2018, she was a Visiting Student with the Photogrammetry and Image Analysis Department, Remote Sensing Technology Institute (IMF), German Aerospace Center (DLR), Oberpfaffenhofen, Germany. She is currently a Postdoctoral Researcher with the School of Information and Electronics, BIT. Her research interests include signal/image processing, fractional Fourier transform, deep learning, and their applications in biometrics and geospatial object detection.

Dr. Wu was a recipient of the Jose Bioucas Dias Award for the outstanding paper at the Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing (WHISPERS) in 2021.

Jing Yao received the B.Sc. degree from Northwest University, Xi'an, China, in 2014, and the Ph.D. degree from the School of Mathematics and Statistics, Xi'an Jiaotong University, Xi'an, China, in 2021.

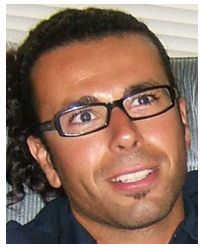
He is currently an Assistant Professor with the Key Laboratory of Digital Earth Science, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, China. From 2019 to 2020, he was a Visiting Student with Signal Processing in Earth Observation (SiPEO), Technical University of Munich (TUM), Munich, Germany, and with the Remote Sensing Technology Institute (IMF), German Aerospace Center (DLR), Oberpfaffenhofen, Germany. His research interests include low-rank modeling, hyperspectral image analysis, and deep learning-based image processing methods.

Adil Mehmood Khan received the B.S. degree in information technology from the National University of Sciences and Technology (NUST), Islamabad, Pakistan, in 2005, and the M.Sc. and Ph.D. degrees in computer engineering from Kyung Hee University, Seoul, South Korea, in 2011.

He is currently a Professor with the Institute of Artificial Intelligence and Data Science, Innopolis University, Innopolis, Russia. His research interests include machine learning and deep learning.

Manuel Mazzara received the Ph.D. degree in computing science from the University of Bologna, Bologna, Italy, in 2006, cooperated with European and U.S. industry, plus governmental, and intergovernmental organizations such as the United Nations.

He is currently a Professor of Computer Science with Innopolis University, Innopolis, Russia, with a research background in software engineering, service-oriented architectures and programming, concurrency theory, formal methods, software verification, and artificial Intelligence. His research interests include the development of theories, methods, tools, and programs covering the two major aspects of software engineering and artificial intelligence: the process side, describing how we develop software, and the product side, describing the results of this process.



Salvatore Distefano received the M.Eng. degree (Summa Cum Laude) in computer engineering from University of Catania, Italy, in 2001, and the Ph.D. degree (Summa Cum Laude) in computer science and engineering from University of Messina, Italy, in 2006.

He is currently an Associate Professor with the University of Messina, Messina, Italy. He has authored and co-authored more than 250 scientific papers and contributions to international journals, conferences, and books. He has visited as a Scholar and Professor different universities and research centers such as collaborating with top scientists such as UMass Dartmouth, UCLA, Duke, Innopolis, and Kazan Federal University. He took part in several national and international projects, such as Reservoir, Vision (EU FP7), SMSCOM (EU FP7 ERC Advanced Grant), Beacon, and IoT-Open.EU (EU H2020). During his research activity, he contributed to the development of several tools such as WebSPN, ArgoPerformance, GS3, and Stack4Things. He is one of the Co-Founders of the SmartMe.io start-up, a spin-off of the University of Messina established in 2017. His research interests include non-Markovian modeling; quality of service/experience; parallel and distributed computing, grid, cloud, autonomic, volunteer, crowd, edge, fog computing; Internet of Things; cyber-physical social systems; smart cities; intelligent transportation systems; Big Data, stream processing; software-defined and virtualized ecosystems; hyperspectral imaging; and machine learning.

Dr. Distefano is a member of international conference committees and he is on the editorial boards of IEEE TRANSACTIONS ON DEPENDABLE AND SECURE COMPUTING, *Journal of Cloud Computing*, and *International Journal of Big Data*.

Jocelyn Chanussot (Fellow, IEEE) received the M.Sc. degree in electrical engineering from the Grenoble Institute of Technology (Grenoble INP), Grenoble, France, in 1995, and the Ph.D. degree from the Université de Savoie, Annecy, France, in 1998.

Since 1999, he has been with Grenoble INP, where he is currently a Professor of Signal and Image Processing. He was a Visiting Scholar with Stanford University, USA, KTH, Sweden, and NUS, Singapore. Since 2013, he has been an Adjunct Professor with the University of Iceland, Reykjavik, Iceland. From 2015 to 2017, he was a Visiting Professor with the University of California, Los Angeles (UCLA), USA. He is currently the AXA Chair in Remote Sensing and is an Adjunct Professor with the Chinese Academy of Sciences, Aerospace Information Research Institute, Beijing, China. His research interests include image analysis, hyperspectral remote sensing, data fusion, machine learning, and artificial intelligence.

Dr. Chanussot is the Founding President of IEEE Geoscience and Remote Sensing French chapter (2007–2010) which received the 2010 IEEE GRS-S Chapter Excellence Award. He has received multiple outstanding paper awards. He was the Vice-President of the IEEE Geoscience and Remote Sensing Society, in charge of meetings and symposia (2017–2019). He was the General Chair of the first IEEE GRSS Workshop on Hyperspectral Image and Signal Processing, Evolution in Remote sensing (WHISPERS). He was the Chair (2009–2011) and Cochair of the GRS Data Fusion Technical Committee (2005–2008). He was a member of the Machine Learning for Signal Processing Technical Committee of the IEEE Signal Processing Society (2006–2008), and the Program Chair of the IEEE International Workshop on Machine Learning for Signal Processing (2009). He is an Associate Editor for IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, IEEE TRANSACTIONS ON IMAGE PROCESSING, and *Proceedings of the IEEE*. He was the Editor-in-Chief of *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* (2011–2015). In 2014, he was a Guest Editor for the *IEEE Signal Processing Magazine*. He is a member of the Institut Universitaire de France (2012–2017) and a Highly Cited Researcher (Clarivate Analytics/Thomson Reuters).