Conventional to Deep Ensemble Methods for Hyperspectral Image Classification: A Comprehensive Survey

Farhan Ullah, Irfan Ullah, Rehan Ullah Khan, Salabat Khan, Khalil Khan, and Giovanni Pau, *Member, IEEE*

***Abstract*—Hyperspectral image classification (HSIC) has become a hot research topic. Hyperspectral imaging (HSI) has been widely used in a wide range of real-world application areas due to the in-depth spectral information stored within each pixel. Noticeably, the detailed features, i.e., a nonlinear correlation between the obtained spectral data and the correlating HSI data object, generate efficient classification results that are complex for traditional techniques. Deep learning (DL) has recently been validated as an influential feature extractor that efficiently identifies the nonlinear issues that have arisen in various computer vision challenges. This motivates using DL for HSIC, which shows promising results. This survey provides a brief description of DL for HSIC and compares cutting-edge methodologies in the field. We will first summarize the key challenges for HSIC, and then, we will discuss the superiority of DL and DL ensemble in addressing these issues. In this article, we divide state-of-the-art DL methodologies and DL with ensemble into spectral features, spatial features, and combined spatial–spectral features in order to comprehensively and critically evaluate the progress (future research directions as well) of such methodologies for HSIC. Furthermore, we will take into account that DL involves a substantial percentage of labeled training images, whereas obtaining such a number for HSI is time and cost consuming. As a result, this survey describes some methodologies for improving the classification performance of DL techniques, which can serve as future recommendations.**

***Index Terms*—Convolutional neural network (CNN), deep ensemble, deep learning (DL), hyperspectral image classification (HSIC), spatial features, spectral features.**

I. Introduction

HYPERSPECTRAL imaging (HSI) is involved with the retrieval of relevant data relying on the radiance obtained through the optics spectrographic imaging system at short or long ranges without significant interaction with the object of interest [1], [2]. HSI provides rich spectral information by acquiring reflectance measurements across hundreds of narrow contiguous spectral bands over a wide wavelength range from 0.4 to 2.4 μm. This span covers the visible region from 0.4 to 0.7 μm as well as the shortwave infrared region from 0.7 to 2.4 μm. The fine spectral sampling at each spatial location enables detailed characterization of the spectral signature for all materials in a scene. In this way, HSI offers comprehensive spectroscopic data that facilitate the identification and discrimination of materials that may appear similar in broad multispectral bands. The combination of high spectral resolution and broad wavelength coverage makes HSI a versatile tool for detailed material detection and analysis. HSI is also capable of investigating the (light) emission features of instances in the mid- to long-infrared range [3].

Alongside the comprehensive information, HSI poses some issues considering conventional analysis methods for monochromatic, RGB, and multispectral images that are unable to explicitly leverage hyperspectral images to retrieve relevant data for a variety of factors, e.g., high-dimensional spectral and spatial information (HSI) displays the distinctive statistical and geometrical features, i.e., in a hypercube and hypersphere, the volume is primarily focused on the edges and outer shells, respectively.

HSI has been used in a variety of real-world applications, such as atmosphere, climate, urban, agricultural production, geomorphological and mineral investigation, coastal area, maritime, forest management, groundwater and surface contamination, waterways, glaciers and ice sheets, biological, health-care cases, and food computation [4], [5], [6], [7], [8], [9]. There are also military service uses such as camouflage, landmine detection, and coastal defense area navigation. Moreover, HSI has been employed in space, air, and underwater vehicles to obtain comprehensive spectral data for a variety of applications [10], [11], [12], [13].

For any of the aforementioned application domains, infield sampling and spectral library indexing of ground truth samples are crucial for various factors. For example, the spectral data of grasslands are hampered by a variety of ecological conditions, which render it difficult to portray variance sufficiently without the gathering of location-specific ground spectra. However, the true significance of HSI is relatively unexplored because it

Manuscript received 1 November 2023; revised 26 December 2023; accepted 31 December 2023. Date of publication 12 January 2024; date of current version 1 February 2024. This work was supported by Kore University of Enna and IEEE mutual agreement. *(Corresponding author: Giovanni Pau.)*

Farhan Ullah and Salabat Khan are with the College of Computer Science and Software Engineering, Shenzhen University, Shenzhen 518060, China (e-mail: farhan.marwat@gmail.com; salabatwazir@gmail.com).

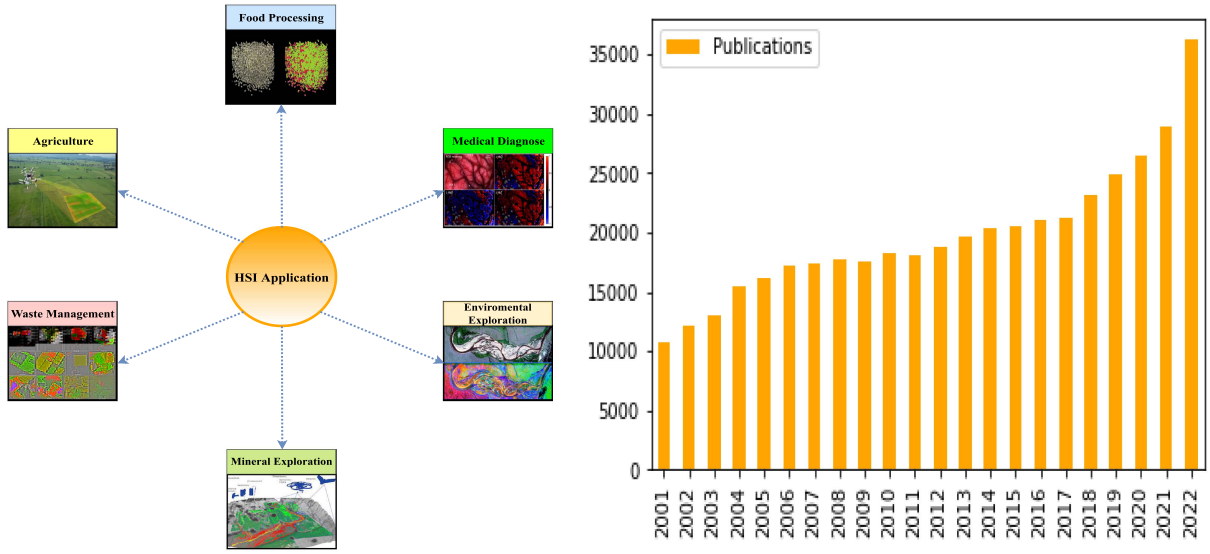
Irfan Ullah is with the School of Computer Science, Chengdu University of Technology, Chengdu 610059, China (e-mail: irfan.ee@cqu.edu.cn).

Rehan Ullah Khan is with the Department of Information Technology, College of Computer, Qassim University, Buraydah 52571, Saudi Arabia (e-mail: re.khan@qu.edu.sa).

Khalil Khan is with the Department of Computer Science, School of Engineering and Digital Sciences, Nazarbayev University, Astana 010000, Kazakhstan (e-mail: khalil.khan@nu.edu.kz).

Giovanni Pau is with the Kore University of Enna, 94100 Enna, Italy (e-mail: giovanni.pau@unikore.it).

Digital Object Identifier 10.1109/JSTARS.2024.3353551



(a) (b)

Fig. 1. (a) Various real-world applications of HSI and (b) remote sensing/HSIC-relevant research papers published per year till April 25, 2023. [Source: Google Scholar accessed on April 25, 2023, and the total number of research papers (including patents and citations) were classified by its relevancy.]

enables it to penetrate greater depth than surface features, even though each feature typically has a distinct spectrum band. As such, HSI can acquire over 200 spectral bands, allowing experts to allow discriminate objects that were previously impossible. Fig. 1 depicts some of the HSI application scenarios. Moreover, a number of other fields, for example, smart cities, smart industrial processing, and intelligent transportation systems, can significantly help using this methodology.

Given the limitations mentioned above, HSI evaluation is divided into the following categories: dimensionality reduction (DR) [14], [15], [16], [17], [18] has enabled more efficient processing of hyperspectral data by extracting the most meaningful features and reducing redundant information. Spectral unmixing [19], [20], [21], [22], [23], [24], [25], [26], [27] has decomposed mixed pixels to identify constituent end members and quantify their abundances. This has proved useful for analyzing complex scenes with subpixel composition. Object and change detection algorithms [28], [29], [30], [31], [32], [33], [34] have utilized spectral–spatial approaches to identify areas of interest and track their evolution over time. Classification techniques [35], [36], [37] incorporating spatial context have improved mapping accuracy compared to pixelwise classifiers; feature learning for classification [38], [39], [40], [41], [42] has also been advanced through spectral–spatial feature learning. Furthermore, restoration and denoising methods [43], [44] have integrated spatial regularization to suppress noise while preserving image details. Finally, resolution enhancement approaches [45], [46] have employed spectral–spatial processing to increase the resolution of hyperspectral data. Fig. 1(b) depicts a rapidly increasing pattern in the amount of HSI analysis-related publications each year.

This article primarily focuses on hyperspectral image classification (HSIC), which has attracted the scientific community's attention because of its numerous applications in the fields of land usage and land cover [47], [48], [49], [50], [51], environmental surveillance, hazard identification [52], [53] vegetation mapping [54], [55], and urban planning. To complete the classification task, HSIC techniques use machine learning techniques [56], [57], [58]. These techniques are described in numerous in-depth reviews that were published over the past decade [35], [59], [60], [61], [62], [63], [64], [65], [66], [67]. However, ongoing progress in the area of machine learning occasionally offers better methodologies. One of these breakthrough developments in machine learning that increased HSIC performance was the use of deep learning (DL) approaches [68], [69], [70].

The goal of this article is to provide a synopsis of the popular DL-based HSIC approaches. To be more precise, we will first list the key significant problems with HSIC that conventional machine learning (CML) methods cannot efficiently address, and then, we will list the benefits of DL and DL with ensemble to address these problems. Later on, we will present an architecture to describe the related works into the following groups:

1) spectral and spatial feature learning, individually;
2) reviewing the progress in DL-based HSIC comprehensively using spectral–spatial feature learning;
3) future research studies will aim to optimize the generalization efficiency and durability of DL methods while taking into account the insufficiency of reliable training instances;
4) ensemble-based methods and future recommendations to enhance the performance and improve the time complexity of ensemble-based methods.

II. Background and Challenges

A. Traditional to DL Models

HSIC's primary function is to provide each pixel vector of the HSI cube with a distinct label according to its spectral or spatial features. An HSI cube can be described mathematically as

$$c = [c_1, c_2, c_3, ..., c_d,]\, T \in R_d \times (P \times Q) \quad (1)$$

where d denotes the total number of spectral bands, with each band containing $(P \times Q)$ samples per band belonging to Z classes, where $c_i = [c_{1_i}, c_{2_i}, c_{3_i}, ..., c_{n_i}]T$ is the ith sample in the HSI cube with class label $y_i \in R^y$. Hyperspectral classification aims to learn a mapping function $f_v(.)$ that takes the input data c and generates predicted labels Y by applying transformations to minimize the discrepancy between the predicted and true labels [71]. This constitutes an optimization problem, where $f_v(.)$ maps the high-dimensional input data to the label space by exploiting the rich spectral information and inherent class structure to achieve accurate classification

$$Z = f_v(U, \theta) \tag{2}$$

where θ is a customizable parameter that may be needed to perform transformations on input data v such that $f_v : v \to Z$.

The techniques employed in the computer vision field appeared to have impacted the majority of HSIC frameworks [72]. CML-based HSIC techniques train the classifier with handcrafted features. These techniques are typically based on technical expertise and domain expertise to generate a variety of human-engineered features, such as shape, texture, color, shape, spectral, and spatial descriptions. Each of these features is a fundamental key attribute of an image feature and contains useful data for HSI. Many handcrafted-based feature extraction (FE) and classification methods in HSI have utilized various texture descriptors, including local binary patterns (LBPs) [73], histogram of oriented gradients (HOG) [74], global image scale-invariant transform/global invariant scalable transform (GIST) [75], pyramid HOGs, and scale-invariant feature transform (SIFT) [76]. For the classification stage, techniques such as random forests (RF) [77], kernel-based support vector machine (SVM) [78], K-nearest neighbors (KNN), and extreme learning machines (ELM) have been employed. These FE methods characterize spatial patterns in a scale- and rotation-invariant manner to better discriminate different textures and materials. The classifiers then leverage these descriptive features to categorize hyperspectral pixels or segments into their respective land cover classes. This combination of robust FE and discriminative classification has been shown to enhance mapping accuracy compared to pixelwise spectral approaches alone.

Color histograms are simplistic and efficacious handcrafted-based features that work well for the task of image classification. They are simple to compute and unaffected by minor image alterations such as translation and rotation. A color histogram's primary flaw is that it lacks spatial contextual information, making it challenging to differentiate between instances classes with similar color but unique distributions. Color histograms, which capture the distribution of pixel colors in an image, are prone to variations under changing illumination conditions. In contrast, HOG features provide an alternative descriptor more robust to lighting changes. HOG characterizes the distribution of edge directions within spatial subregions of the image. By compiling histograms of gradient orientations across localized cells, HOG effectively captures the dominant edge structures in a manner invariant to shifts in illumination. Thus, HOG encoding spatial patterns through localized edge orientation histograms can complement color histogram representations, providing improved illumination invariance. The fusion of complementary color and gradient-based features can yield a more distinctive and robust descriptor for image analysis tasks. It has been used in numerous remote-sensing-related research tasks because of its effectiveness in extracting edge and local shape information [47], [79], [80], [81].

SIFT is a widely applied vigorous feature descriptor used for computer vision applications [82], [83], [84], [85]. The SIFT descriptor has the benefit of being insensitive by alteration in image scale, rotation, illumination, and noise. SIFT's drawback is that it is computationally intensive, which raises the computational complexity issue. The scales and orientations (gradient features) of distinct subregions of an image are characterized by GIST, which is the global description of essential factors of an image. GIST generates a spatial frame based on various statistical characteristics, such as roughness, openness, ruggedness, etc. [86]. Texture descriptors such as LBPs are applied for remote sensing image analysis [73], [87]. By selecting pixels from the square neighborhood, LBPs are employed to define the texture around each pixel. The grayscale scores of all neighboring pixels are threshold concerning the central pixel.

The color histograms, GIST descriptors, and various texture representations extract global image features that characterize statistical properties such as color distributions, textural patterns [88], [89], and overall spatial structure [75]. In contrast, local feature descriptors like HOG and SIFT capture geometrical information from local neighborhoods. These local descriptors are often aggregated into bag-of-visual-words (BoVW) models [49], [53], [85], [90], [91], [92], [93], [94], [95] and HOG-feature-based models [47], [96]. Enhancing both the discriminative capability and compactness of BoVW representations has been pursued through various techniques, such as Fisher vector coding [73], [97], [98], spatial pyramid matching [99], and probabilistic topic models [95], [100], [102]. To improve the discriminative power and compactness of BoVW representations, techniques including Fisher vector coding [73], [97], [98], spatial pyramid matching [99], and probabilistic topic model [95], [100], [101], [102]. An integration of such features is utilized in image classification since one feature cannot fully capture the information contained in an image [48], [90], [100], [102], [103], [104], [105], [106], [107], [108], [109].

The diverse attributes of an image can be efficiently defined by handcrafted features, which makes them compatible with the data being explored. Even though these features might not be significant in the scenario of actual data, it can be challenging to fine-tune between rigidity and discriminability because a pattern of ideal features can differ significantly depending on the type of data. In addition, since handcrafted features must be designed with a high level of domain skill, human assistance in their design has a significant impact and effectiveness during the classification stage.

To alleviate the constraints of the handcrafted feature process, a deep feature learning approach was presented by Hinton and Salakhutdinov [110]. DL-based methodologies are applied to extract features from data in a hierarchical fashion to build a model with progressively increasing semantic layers. This is accomplished once the model acquires a suitable representation of the data. Such models have demonstrated promising effects for feature representation in HSIC [111], [112]. Recently, Li

et al. [113] have introduced a novel baseline network for HAD, named LRR-Net, which combines the LRR model with DL methods. LRR-Net efficiently employs the alternating direction method of multipliers optimizer to address the LRR model and integrates the obtained solution as prior knowledge into the deep network. This integration guides the optimization of parameters effectively.

DL frameworks can extract the patterns of any data without any preliminary information on the statistical distribution of input data [114]. Furthermore, these frameworks can capture linear and nonlinear features from the input data without any preexisting information. Such frameworks can handle HSI data in both the spectral and spatial aspects independently, as well as in an associated manner. DL frameworks have a versatile design in contexts of layer typologies and complexity, and they are adaptable to different machine learning techniques such as supervised, semisupervised, and unsupervised techniques.

III. Hyperspectral Data Features and DL Complexities

DL applications for HSI data still present significant challenges, considering the potency that has already been discussed. The majority of such complexities emerge from the features of HSI data, which include numbers of consecutive narrow spectral channels with exceptionally high spectral resolution but limited spatial resolution, covering the entire electromagnetic spectrum. In addition, there is a scarcity of available datasets. Although pixels with abundant spectral features are helpful for classification tasks, processing challenging data requires a considerable investment of time and resources.

Moreover, because of the higher number of parameters, computing such high-dimensional data entails a hard effort. This is termed the curse of dimensionality because it has a significant impact on classification accuracy, particularly in the context of supervised learning [115]. As described in [116], the Hughes phenomenon occurs when the training examples are limited and unreliable. This means the training data might not offer valuable learning contributions to the model or could contain repetitive patterns. Consequently, effectively training the model becomes challenging, leading to an increased risk of overfitting issues. The phenomenon is particularly noticeable when the amount of labeled training data is significantly smaller than the labeled training data, which is considerably small scale in comparison to the total number of spectral bands available in the data. A key problem in HSIC is the limited number of labeled HSI datasets because labeling HSI typically involves the employment of human experts or the examination of real-world situations, which is a time-consuming and complex operation.

In addition to high dimensionality, HSIC struggles from a variety of additional biases, such as increased intraclass variance produced by unrestrained changes in reflectance values induced by a number of environmental interferers and data deterioration generated by instrumentation distortion during data capture [117]. Similarly, the inclusion of overlapping bands related to HSI instrumentation influences the computational complexity of the model. Another problem associated with HSI spatial resolution is spectral mixing. HSI pixels with low to medium spatial resolution comprise extensive spatial areas of the earth's land, resulting in heterogeneous spectral signatures with considerable interclass consistency in border regions. Consequently, identifying materials relying on their spectrum reflectivity gets challenging [118]. Some major issues that arise when DL is performed on HSIC are listed as follows.

1) *Training complexity:* Deep neural network (DNN) training and optimization via optimizing parameters is an NP-hard issue with no assurance of convergence of the optimization procedure [119], [120], [121]. As a result, it is considered that DNN training is extremely challenging [114], Particularly with HSI, where a huge number of variables must be adapted. Furthermore, as numerous optimization strategies for deep CNNs progress, the process of converging gets somewhat convenient. Among stochastic gradient descent [122] and its momentum version [123], RMSProp [124], Adam [125], AdamW [126], diffGrad [127], RAdam [128], gradient centralization [129], and AngularGrad [130], are effective CNN optimization approaches that are commonly employed in classification tasks.
2) *Training data scarcity issue:* As previously stated, a supervised DNN needs a significant size of training samples; otherwise, its inclination to overfit grows substantially, resulting in the Hughes phenomenon [131]. The high dimensionality of HSI, combined with a limited set of labeled training samples, renders DNNs inefficient for HSIC, as it necessitates several tweaks throughout the training process [71].
3) *Model's interpretability:* The DNN training strategy is complex to analyze and comprehend. The black box aspect of DNNs is regarded as a possible vulnerability and may influence the optimization approach architecture procedure. However, considerable progress has been made in interpreting the model's fundamental complexities.
4) *Significant computational liability:* One of the key issues of DNN is interacting with large dataset size, which requires increasing memory bandwidth, significant computational complexity, and storage utilization [132]. Moreover, state-of-the-art approaches, such as parallel and distributed architectures [133], [134] and high-performance computing (HPC) [118], allow DNNs capable of processing enormous volumes of information.
5) *Declining training performance:* It is considered that the DNN retrieves increasingly rich information from data [135]; however, this is not valid for all algorithms to get improved efficiency by merely adding extra layers since enhancing the DNN's levels of depth amplifies the issue of exploding or vanishing gradients [136] and impacts model convergence [135].

IV. HSI Representation

Hyperspectral data are expressed as a 3-D hypercube, $x \in R_B(N \times M)$, which includes data of 1-D spectral and 2-D spatial information, where B denotes the total number of spectral

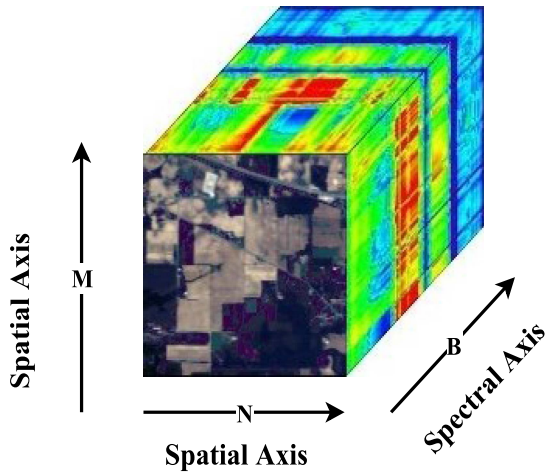


Fig. 2. Hyperspectral spatial and spectral cube.

bands and N and M represent spatial features, i.e., width and height, respectively. The HSI cube is exhibited in Fig. 2.

A. Spectral Representation

In these expressions, each pixel vector is derived from surrounding pixels and deeply analyzed solely on spectral characteristics. This denotes that the pixel is expressed exclusively in the spectral space, denoted as $x_i \in R_B$, where B can be the total number of spectral channels or just salient spectral bands retrieved via DR technique. In order to eliminate redundancies and obtain greater class feature subsets without significantly losing important information, a low-dimensional description of HSI is typically recommended for data analysis rather than employing the actual spectral bands.

There are two types of DR techniques for spectral HSI interpretation: supervised and unsupervised. DR techniques like principal component analysis (PCA) and locally linear embedding provide unsupervised methods to transform high-dimensional hyperspectral data into a lower dimensional space. By projecting the data onto a set of orthogonal principal components (PCs) or through locally linear reconstructions, these techniques can extract the most salient features and inherent data structure without needing prior class labels [137]. The reduced representation improves computational and statistical efficiency for subsequent analysis while concentrating the meaningful information content. Thus, unsupervised DR offers an effective approach for exploiting the rich spectral data in hyperspectral imagery in an exploratory manner, prior to employing supervised classification or detection algorithms requiring annotated training samples.

In contrast, supervised DR approaches use labeled samples to learn the data distribution. This involves ensuring that data points from similar categories are clustered closely together while simultaneously separating data points from various categories. For example, linear discriminant analysis (LDA), local Fisher discriminant analysis (LFDA) [138], local discriminant embedding [139], and nonparametric weighted FE [140]. LDA and LFDA strengthen class feature space by increasing data point interclass proximity and decreasing intraclass proximity. However, because of the spectral mixing phenomenon, which causes identical information to occur with multiple spectra or distinct information to have similar spectral fingerprints, it is hard to distinguish between various classes solely relying on spectral reflectance scores.

B. Spatial Representation

While spectral signatures provide valuable information, they have limitations in capturing the spatial context in HSI. An alternative approach represents the spatial arrangement by expressing each band as a matrix, $x_i \in R^{N\times M}$, containing the pixel values. Given strong spatial correlations, neighboring pixels likely share class labels. Local pixel neighborhoods can be incorporated with spatial representation, computing a pixel's locality within a window [141]. Morphological profiles (MPs), various texture features, such as Gabor filters (GFs), gray-level co-occurrence matrix (GLCM), LBPs, etc., and DNN-based algorithms are some popular techniques for extracting spatial information from HSI cubes. Geometrical properties can be extracted from MPs. Some improvement in MPs involves extended morphological profiles (EMPs) [142], multiple-structure-element MPs [143], and invariant attribute profiles [144].

The image processing technique produces valuable spatial contextual features for HSI. For example, a texture analysis tool called GF can effectively gather textural information at different spatial and temporal scales. Consequently, LBP can give rotation-invariant spatial texture representation. The spatial variance of HSI can be efficiently calculated by the GLCM, which considers the respective positions of neighbor pixels. DNNs, on the other hand, can retrieve spatial features from HSI by treating individual pixels as image patches instead of spectral vectors. In addition, the spatial features included in HSI can also be retrieved by merging previously mentioned approaches. For example, Zhang et al. [145] merged the GF method and differential morphological profiles (DMPs) method [146] to retrieve local spatial sequential (LSS) features for a recurrent neural network (RNN)-based HSIC architecture.

C. Spectral–Spatial Representation

This representation uses both spectral and spatial data features. In these kinds of methods, a pixel vector is evaluated by using its spectral properties while taking spatial contextual features into account. Approaches that employ combined spectral and spatial representations of HSI integrate spatial features with spectral vectors [64], [147] either to deeply execute the 3-D HSI cube or to maintain the true topology and contextual features [148].

All of these HSI representations are frequently used for HSIC in the literature. The spectral representation of HSIs was used by the majority of DNNs for pixel-level classification [149], [150]. Although numerous attempts have been performed to include spatial information in order to alleviate the shortcomings of spectral representation [151], [152]. A recent trend that has gained considerable attention and enhanced classification performance is the simultaneous exploration of spectral and spatial data [14], [69], [153], [154], [155], [156]. In Section V, there is a more

in-depth discussion of the HSI feature exploration methods for HSIC.

V. Learning Strategies

DL models can use a variety of learning approaches, which can be widely classified as follows.

A. Supervised Learning Approach

The supervised learning approach involves training a model using a labeled dataset that consists of input data along with their respective class labels. In the training step, the model adjusts its hyperparameters to predict the intended outcomes precisely. In the testing step, the model is evaluated using raw test data to confirm its accuracy in predicting the right labels. In order to train the model properly, it can make predictions of the labels of raw input data. Despite the advantages of the supervised learning approach in DNNs, it necessitates a substantial number of labeled training datasets to effectively adjust or optimize the model hyperparameters [58], [157], [158]. As a result, they are particularly fit for cases with a huge size of labeled data. The specifics of different supervised learning strategies for DNNs will be covered in the following sections.

B. Unsupervised Learning Approach

Unlike the supervised learning approach, unsupervised learning methods extract features from input data that do not have explicit labels. In the lack of relevant labels, these techniques attempt to determine the fundamental statistical pattern of input representations. The lack of ground truth samples for the training dataset can make evaluating the effectiveness of the trained model challenging. However, such learning algorithms are effective when we aim to extract the intrinsic pattern of such datasets with limited training samples. PCA is an unsupervised learning method generally employed to generate a low-dimensional representation of the input data. Likewise, the k-means clustering method is an unsupervised learning technique that generates different categories of input data into unified groups.

C. Semisupervised Learning Approach

Semisupervised learning (SSL) is the third approach that falls between unsupervised and supervised methodologies. SSL techniques learn from partly labeled dataset, i.e., a limited amount of labeled training data can be used to label the remaining unlabeled data. Since such strategies efficiently employ entire available data rather than only labeled data, they have received wide attention in the scientific community and are frequently utilized in HSIC [159], [160], [161], [162]. In Section XI, the specifics of these techniques are briefly discussed.

VI. DNN Construction (Layer Composition)

In this section, we will examine the latest advances in some popular DNN frameworks for HSIC. We primarily reviewed the publications released after 2017. DNNs offer a wide range of versatile and scalable HSIC models that enable the integration of various sorts of layers. The following sections cover a few frequently employed sorts of layers.

A layer is a fundamental structural component of DNN, and the sort of layer has a significant effect on data processing. A layer accepts the weighted input, transforms it linearly or nonlinearly, and transfers the values to the next layer. A layer usually has a singular activation function, making it uniform. The input layer marks the initial layer of the network, and the output layer represents the final layer. The layers that lie between the input and output layers are commonly known as hidden layers. These layers conduct diverse operations to extract various features in the input data. The layered form chosen is determined by the problem, as certain layers behave more effectively than others. The most widely employed HSIC layers are described as follows.

A. Fully Connected Layers

Fully connected (FC) layers integrate each individual neuron in the bottom layer into every neuron in the upper, creating dense connections between the layers and enabling the propagation of information throughout the neural network. Following convolution/pooling layers, they are typically employed as the final few layers in a model. In predicting the probability for class labels, FC leverages the output of the preceding layer. The computational complexity is significantly increased because there are many interconnections and, hence, many hyperparameters that need to be tweaked. In addition, the model is highly prone to overfitting problems due to the significant number of hyperparameters [50]. Therefore, a dropout approach is adopted in order to lessen the impact of the overfitting issue [163].

B. Convolutional Layers

In HSIC, convolutional layers are a fundamental component of convolutional neural networks (CNNs) designed to process hyperspectral data [164]. These layers are specifically tailored to handle the unique characteristics of hyperspectral images, where each pixel contains information across multiple spectral bands. An explanation of convolutional layers in the context of HSIC is given as follows:

1) *Input representation:*
   a) *3-D Tensor:* Hyperspectral images are represented as 3-D tensors, where each pixel contains a spectrum of information across multiple bands. The dimensions of the tensor are (height, width, bands). For example, a 100×100 pixel hyperspectral image with 100 spectral bands would have dimensions $100 \times 100 \times 100$.
2) *Convolutional operations:*
   a) *Spatial convolution:* Traditional 2-D convolutional operations are applied to capture spatial patterns and relationships between pixels. Filters are used to detect features in the spatial domain.
   b) *Spectral convolution:* 1×1 convolutions are often employed along the spectral dimension to capture spectral dependencies. These convolutions allow the network to learn combinations of spectral bands, enhancing its ability to extract relevant information.

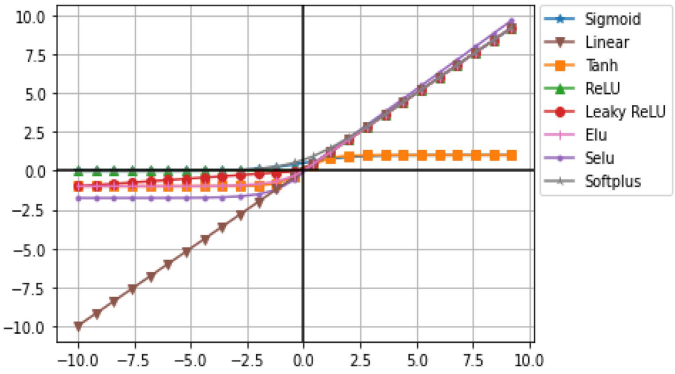


Fig. 3. Graphical depiction of different frequently employed activation functions.

3) *Feature extraction:*
   a) *Learned filters:* Convolutional layers use learned filters to convolve over the input data. These filters act as feature detectors, capturing local patterns in both spatial and spectral dimensions.
4) *Hierarchical feature extraction:*
   a) *Multiple convolutional layers:* Convolutional layers are often stacked to form deep architectures, allowing the network to learn hierarchical features from both spatial and spectral information. Each subsequent layer captures increasingly abstract and complex features.
5) *Parameter sharing:*
   a) *Shared weights:* Convolutional layers use shared weights across the input space, reducing the number of parameters compared to FC layers. This parameter sharing enhances the model's ability to generalize to different spatial locations and spectral bands.

C. Activation Layers

In DNNs, the activation layers are considered the feature detector phase [165] because they play a crucial role in capturing and learning relevant patterns and representations from the input data. The FC and convolutional layers either produce linear representations of input data or treat them in a manner analogous to linear regressors, and the data that undergo transformation by these layers are described as being in the FE phase [69]. As a result, an activation layer must be utilized after the FC layer and convolutional layers to retrieve the nonlinear characteristics of data. Feature maps from earlier layers are processed via an activation function that is applied to generate an activation map within the activation layer. Sigmoid, hyperbolic tangent (tanh), rectified linear unit (ReLU), LiSHT [166], and SoftMax are a few examples of regularly applied activation functions. The activation functions SoftMax and ReLU are frequently used in HSI analysis [71]. A few frequently used activation functions are shown graphically in Fig. 3.

D. Pooling Layers

The pooling layers are also referred to as downsampling layers, which accept a given input data size and transform it into a single value, as illustrated in Fig. 4. This enables invariance to minimal data anomalies. Since the volume of the input and the model hyperparameters are both minimized, which also reduce the computing time, the pooling layer aids the model in preventing overfitting problems. Among various downsampling techniques, three most popular ones include max pooling, average pooling, and sum pooling. Furthermore, a new optimized pooling method called wavelet pooling was presented in [167], and its effectiveness is comparable to that of max pooling and average pooling. Similarly, Springenberg et al. [168] presented an alternative approach in which the convolutional layer of enhanced filter stride takes the role of the pooling layer.

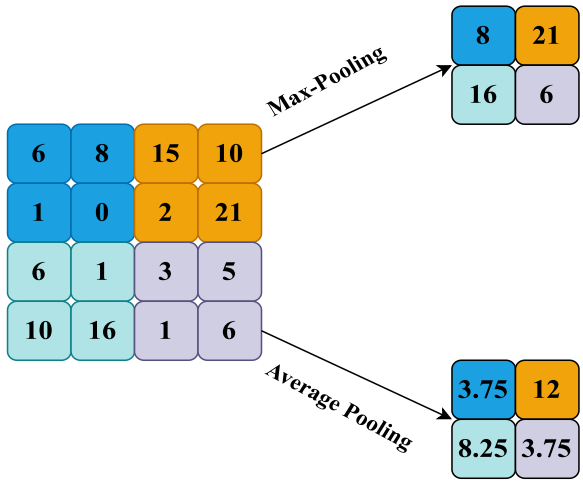


Fig. 4. Graphical depictions of the max-pooling and average-pooling procedures of the downsampling/pooling layer.

VII. Convolutional Neural Network

The CNN architecture draws inspiration from the biological visual system, as described in [169]. Adopting Hubel and Wiesel's [169] natural visual recognition process, Neocognitron [170] is recognized as the original hierarchical position-invariant pattern recognition model [171] and can be regarded as the prior of CNN [172]. The CNN architecture is categorized into two phases: FE network and classification according to the feature maps generated during the initial step.

The FE network is made up of several hierarchical structures positioned in convolutional, activation, and pooling layers. The convolutional layer retrieves features from input data by convoluting it with a learned kernel. The complexities of the model are minimized, and the network is more straightforward to train as a result of the decrease in the number of hyperparameters required to be adjusted by the spatial sharing of the kernel with the entire input data on each convolutional layer. The complexities of the model are minimized, and the network is simpler to train as a result of the decrease in the number of hyperparameters required to be tweaked by the spatial sharing of the kernel with the entire input data on each convolutional layer. In order to retrieve nonlinear features from the input, convolution outputs are then sent via an activation layer that contributes nonlinearities to the network. This is accomplished by taking the outputs of

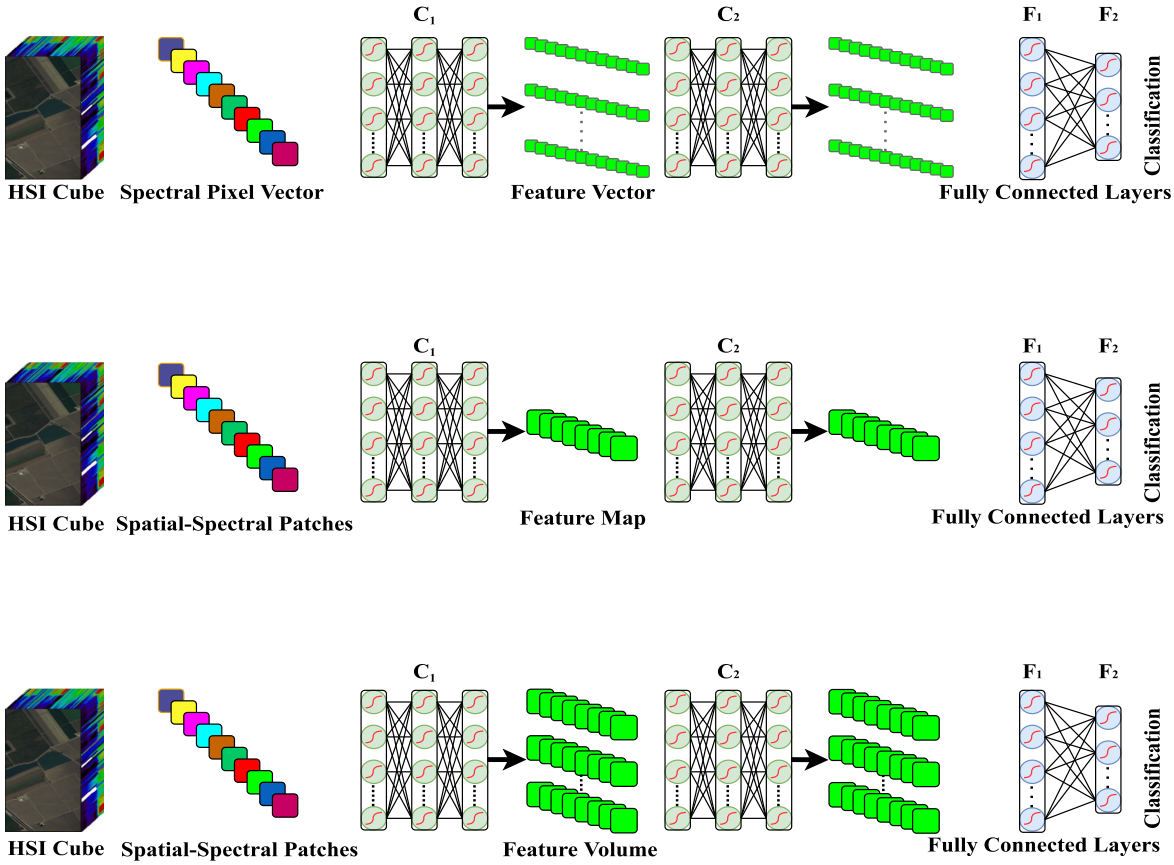


Fig. 5. Generic structure of spectral CNN, spatial CNN, and spectral–spatial CNN frameworks for HSIC.

the convolution and adding a nonlinear function to them. After that, a pooling process is employed to lower the feature map's resolution in order to obtain shift invariance. In basic terms, the pooling layer is placed along each convolutional layer, preceded by the activation function.

Given the feature maps generated at the FE step, a SoftMax function and FC layers are used in the classification phase to calculate the likelihood that an input pattern belongs to a given class. Each neuron in the last layer is linked to every neuron in the current layer by the FC layer. The authors of [173] and [174] presented that the FC layer can be skipped by employing a global average pooling (GAP) layer. SoftMax is frequently employed for classification problems [175], [176], although SVM [177], [178] has also been employed for this objective in numerous pieces of research.

In the next sections, we will discuss three distinct categories of CNN structures for HSIC: 1) spectral CNN frameworks; 2) spatial CNN frameworks; and 3) spectral–spatial CNN frameworks. Fig. 5 shows the overall structure of these three approaches.

A. *Spectral CNN Frameworks for HSIC*

Spectral CNN models only receive 1-D spectral information as input ($x_i \in R^B$), where B could be the actual number of spectral bands or the appropriate number of bands retrieved after the DR method. In the research study [179], a CNN architecture was presented to reduce the overfitting issue and acquire improved generalization ability via using the 1×1 convolutional kernels and increased dropout rates. Furthermore, the incorporation of a GAP layer in lieu of an FC layer serves a dual purpose in the neural network architecture. It not only effectively reduces the overall number of network hyperparameters but also brings about a notable enhancement in the model's ability to generalize and avoid overfitting. Gao et al. [174] introduced a CNN structure for HSIC that thoroughly exploited the spectral data by incorporating the 1-D spectral vector to a 2-D feature matrix, and the architecture obtained feature reusability potential by cascading fusion of layers composed of 1×1 and 3×3 convolutional layers. They [174] used the GAP layer in a similar way [179] to reduce the total number of the network's training parameters optimized during the training process, allowing it to extract high-dimensional features from the data.

A hybrid model for HSIC was proposed in [180], in which the initial convolutional layers are applied to retrieve position-invariant middle-level features and then employing recurrent layers are utilized to retrieve spectral-contextual information. Consequently, Jin et al. [149] classified healthy and diseased wheat heads using a hybrid architecture. They incorporate spectral data into a 2-D data structure for the input layer. In [181], it was shown that the CNN performed better than SVM and KNN for spectral-based variation detection of rice seeds. A related CNN application was investigated in [150], where different Chrysanthemum types were recognized utilizing spectral information from the first five PCs of PCA. A common DR

technique for handling/preprocessing high-dimensional data in DL applications is PCA. In [182], PCA was used to preprocess medical HSI before CNN kernels and Gabor kernels were fused with dot products to perform classification.

Charmisha et al. [183] investigated another DR method called dynamic mode decomposition, which transforms 3-D HSI data to 2-D before feeding it to vectorized CNN for classification. In [184], a technique of averaged spectrum was employed to minimize the noise impact in pixelwise HSIC, where an averaged spectrum of a set of pixels corresponding to bacterial colonies was retrieved for detailed evaluation.

B. Spatial CNN Frameworks for HSIC

Spatial CNN frameworks mainly incorporate spatial features and are also applied to retrieve spatial features from HSI data; spectral-domain DR approaches are used to minimize the complexity of the actual HSI data. For example, Li et al. [185] extracted the first PC enriched with enhanced spatial features. They then fed these features into a full CNN architecture for a classification task. Moreover, Haut et al. [186] used a single PC to train a spatially based 2-D CNN. In [187], the input data preprocessed using PCA whitening and limited to three PCs are fed into a randomized patches network, which serves as a 2-D CNN classification architecture. A limited number of training instances with spectrally comparable features can make deep learning (DL) models susceptible to overfitting issues. To mitigate this problem, Wang et al. [189] introduced an innovative solution named Probabilistic Neighborhood Pooling-based Attention Network (PNPAN), specifically designed for HSIC.

The approach presented in study [189] exploited patches clipped from 2-D input instances (i.e., images from various spectral bands) to train a 2-D CNN framework that acquires the data-adaptive kernels on its own. Moreover, several authors presented using handcrafted features in conjunction with spectral-domain minimization. For instance, to avoid the overfitting issue caused by insufficient training data, Chen et al. [190] integrated the Gabor filtering method with the 2-D CNN for HSIC. Gabor filtering captures spatial features such as edges and textures, significantly reducing the overfitting issue. The study in [191] presented a flexible HSIC network based on the idea of flexible sampling locations that can vary their size and shape to the spatial properties of the HSI. These sampling locations are generated by computing 2-D offsets for each pixel in the input image using standard convolutions and three PCs. These offsets can contain the locations of adjacent nearby pixels with corresponding properties. The structural features of nearby pixels are further combined to produce flexible feature images. Standard convolution on such flexible feature images can retrieve better efficient complicated structures.

C. Spectral–Spatial CNN Frameworks for HSIC

In terms of pixels, spectral–spatial HSIC can be acquired by combining spectral data with spatial features. For example, Ran et al. [192] introduced a better pixel pair feature (PPF) method referred to as spatial PPF. This method differs from conventional PPFs in two key ways: first, the choice of the pixel pair; only pixels from the instantaneous vicinity of the central pixel can further be utilized to form a set; and second, the label of the pixel pair would be as of the central pixel. In another study, Zhong et al. [193] proposed a supervised spectral–spatial residual network that employs a series of 3-D convolutions in the appropriate spectral and spatial residual blocks to retrieve discriminative joint representation. An effective deep 3-D CNN approach that effectively uses both spectral and spatial data for HSIC was put forth in [194].

Recent work has developed advanced DL techniques to extract informative features from hyperspectral images while modeling spatial context. Li et al. [195] employed an adaptive weight learning approach to represent variances across spatial neighborhoods in hyperspectral patches. In addition, Roy et al. [156] proposed a flexible framework with adjustable receptive fields, along with an enhanced spectral residual network for joint FE. This creates more versatile convolutional kernels. As studied in [196], combining max and min convolutional features before the ReLU activation can further improve the discriminative capability of extracted features for classification. Paoletti et al. [197] presented the translation-equivariant interpretations of input features, which give the spatial feature locations for HSIC additional durability because CNNs were unable to leverage rotation equivariance normally.

Given the limited number of training instances that can be labeled, DNNs can also experience overfitting and gradient vanishing issues. To address this issue, lightweight CNNs are gaining increasing interest in HSIC research societies. An end-to-end 3-D lightweight CNN was presented in [198] to solve the problem of the HSIC's scarcity of training instances. Jia et al. [199] introduced a novel approach to extracting the spatial–spectral features by Schrödinger eigenmap joint spatial–spectral features and gradually reducing the dimensionality by employing the compression method to bridge the significant disparity between the huge learnable parameters and the sparsely labeled instances. An estimated 90% of the overall trainable weights are utilized instantaneously after the flattening process, i.e., in the FC layer, with the residual 10% employed on the network's prior convolutional layers. To tackle the challenge of joint spectral–spatial feature learning, Roy et al. [200] presented a novel end-to-end framework incorporating lightweight bag-of-feature learning into a spectral–spatial squeeze-and-excitation residual network architecture for HSIC. This synergistically combines an attention-based residual network that extracts discriminative features from spectral and spatial dimensions with a bag-of-features module that aggregates contextual information from local neighborhoods. By learning an integrated spectral–spatial representation in an end-to-end manner, this method achieves strong performance for hyperspectral scene understanding.

Morphological approaches, such as erosion and dilation, are effective nonlinear feature transitions extensively utilized to maintain an image's vital shape and structural features. In response to such findings, Roy et al. [201] proposed a novel morphological framework called end-to-end morphological CNN for HSIC, integrating both spectral and spatial information through hybridizing the outputs of spectral and spatial morphological blocks retrieved in a dual-path framework.

Li et al. [195] presented a duplet-phase approach for joint spectral–spatial HSIC that can effectively generate spectral and spatial features rather than hybridizing them separately. The first phase of the presented approach consists of a CNN and SoftMax normalization that retrieves joint shallow features and learns the input patch weights in an improved manner. Once deep hierarchical features have been obtained from these shallow features via a network of stacked autoencoders (SAEs), a multinomial logistic regression (MLR) layer is applied to process the last step of classification. In order to jointly leverage spectral features from HSI, a 3-D CNN model was presented in [202]. To analyze its effectiveness, a comparative analysis is made with a spectral-based deep belief network (DBN), an SAE, and a 2-D spatial CNN for HSIC. Roy et al. [203] presented a novel fused squeeze-and-excitation network, which is a bilinear fusion methodology, and the two squeeze procedure parts premised on the global and max-pooling process, while the excitation process is carried out using the squeeze operation's fused output.

A deep multiscale spectral–spatial FE technique for HSIC that can extract useful discriminant information from samples with high spatial variability was presented in [204]. In the approach, the deep spatial feature is extracted employing a fully convolutional network (FCN), and then, this information is concatenated with spectral data utilizing a weighted fusion approach. Subsequently, classification is done pixel-by-pixel on these fused features.

Zhang et al. [205] proposed an efficient and novel dual-channel CNN architecture for spectral–spatial HSIC tasks. The presented method first employed a 1-D CNN to retrieve hierarchical spectral information, and 2-D CNN was applied to retrieve hierarchical spatial information. In the second step, such details are merged for the ultimate classification step. Moreover, to mitigate the insufficiency of training data and obtain better classification performance, the presented approach is supplemented by a data augmentation (DA) method that can boost training instances by a factor of 6. The introduction of a multiscale 3-D deep CNN for end-to-end HSIC in [206] enables the joint learning of both 1-D spectral and 2-D multiscale spatial features without the use of any preprocessing methods like PCA, etc. Dong et al. [207] investigated a novel method for HSIC by integrating a band attention module into the conventional CNN architecture in order to minimize band redundant information in HSI. The finding in [208] introduced an HSIC framework in which multiscale covariance maps that can concurrently extract spectral–spatial information of HSI are employed to acquire multiscale cubes for handcrafted FE by using PCA-transformed images. Following that, the conventional CNN model for classification is trained using these maps.

In [209], the proposed method merged the CNN with the metric-learning-based HSIC approach. In the first phase, the CNN is used to retrieve in-depth spatial features from the initial three PCs retrieved by PCA. In the second phase, metric learning has been incorporated into spectral–spatial feature learning for HSI classification by integrating a metric learning regularization term during model training. This helps increase interclass differences while reducing intraclass variability in the learned feature space. In addition, Gong et al. [210] combine a multiscale convolution-based CNN architecture with diverse deep metric learning techniques based on determinantal point process (DPP) priors [211]. This framework extracts multiscale features across 1-D spectral, 2-D spectral–spatial, and 3-D spectral–spatial modalities. The CNN employs filters at multiple scales to generate descriptive features across scales.

Meanwhile, the DPP-driven metric learning transformer enforces diversity in the feature space to maximize interclass variability while minimizing intraclass variability. By unifying multiscale CNN FE with deep metric learning regularization, this approach achieves robust discrimination through hyperspectral-specific feature optimization, resulting in improved HSI representational tendency. Finally, the last phase of employing the SoftMax classifier has been used to acquire classification maps.

Liu et al. [212] presented an HSIC approach to retrieve multiscale spatial information from HSI by building a three-channel virtual RGB image rather than retrieving the first three PCs via PCA. The goal of employing a three-channel RGB image is to retrieve spatial features incorporating current networks trained on natural images. These images are transmitted to an FCN in order to perform multiscale FE. These multiscale spatial features are combined and then merged with the spectral features retrieved from the PCs for SVM's classification purpose.

Ma et al. [213] proposed a two-branch (spectral and spatial) DNN for HSIC. To retrieve HSI spatial information, the spatial part uses a band selection layer and a convolutional and deconvolutional approach with a skip structure. In contrast, the spectral part uses a contextual DNN to retrieve spectral information. Sellami et al. [214] presented a novel adaptive band-selection-based semisupervised 3-D CNN approach for jointly exploiting spectral information, whereas Roy et al. [215] investigated a dual-attention-based autoencoder (AE)–decoder DNN for unsupervised hyperspectral band selection and consequently merged FE for land cover classification. Correspondingly, spectral–spatial features are utilized in an unsupervised approach in [216] utilizing a 3-D convolutional AE. The inclusion of incorrect/noisy labels in the training sample frequently causes pixelwise land use and land cover (LULC) classification employing conventional CNNs to be overfitted to the labeled noises. To address the issue of efficient classification, Roy et al. [217] presented a novel fragile method of lightweight heterogeneous kernel convolution (HetConv3D) for HSIC with the integration of noisy labels, which efficiently combines spectral and spatial kernel features to generate discriminative and invariant feature maps for classification. Song et al. [218] introduced a deep feature fusion network (DFFN) for HSIC. This network incorporates residual learning to optimize multiple convolutional layers through identity mapping, thereby facilitating the training of deep networks and leveraging increased depth. Consequently, this approach allows the construction of an exceptionally deep network to extract more discriminative features from HSIs.

Roy et al. [219] proposed a hybrid HybridSN framework in which joint spectral–spatial features are initially retrieved using the 3-D CNN, and following the initial FE, the 2-D CNN plays a crucial role in further refining the representation of the data by retrieving extra abstract spatial contextual information. The findings in [220] presented employing adaptive Markov random

field (MRF) for HSIC. Once the CNN has effectively captured joint spectral–spatial features, the next step involves applying smooth MRF priors to the class labels. This strategic utilization of MRF priors allows for the refinement and fine-tuning of the spatial information associated with the predicted class labels. Overfitting and vanishing gradient issues have a significant negative impact on CNN performance. Paoletti et al. [221] proposed the separable attention network, which divides the input feature maps into diverse categories and splits them across the channel dimension before incorporating them to encode the global context data. Lately, Roy et al. [222] have proposed generalized gradient centralized 3-D convolution (G2C-Conv3D) to integrate both intensity-level semantic features and gradient-level in-depth relevant data retrieved from raw HSIs during the convolutional process. G2C-Conv3D can be effectively incorporated into the current HSI FE networks to enhance the effectiveness of effective land types classification. Recently, Song et al. [223] have introduced a novel method, hashing-based deep metric learning, designed for classifying hyperspectral images and light detection and ranging (LiDAR) data. Initially, a two-stream deep network is constructed to independently extract spectral–spatial features from HSI and elevation features from LiDAR.

D. Graph CNN Frameworks for HSIC

Graph convolutional neural networks (GCNNs) [224] are gaining popularity among the research community in a variety of application domains due to their adaptable and diverse network architectural design, which can handle nongrid high-dimensional data. Such characteristics offer new insights and opportunities for more effective and efficient computation of hyperspectral data. GCNNs, in particular, allow for the modeling of data relationships. As a result, we are generally motivated to employ GCNNs to acquire the spatial relationships of spectral signatures in HSIs. Despite GCNNs' constraints in graph construction [225], especially for sizeable graphs (requiring costly computational complexity), GCNNs suffer from classifying materials in massive-scale hyperspectral scenes utilizing standard PCs, resulting in a lower attractiveness in HSIC comparison to CNNs. As a result, there have been several preliminary studies employing GCNNs in the HSIC function.

For instance, a second-order graph convolutional network (SO-GCN), a novel approach to modeling spatial features on manifolds for HSIC [226]. The primary objective of this innovation was to reduce the computational burden associated with graph processing. Adding to this discourse, Wan et al. [227] pioneered the superpixel segmentation method on HSIs, feeding superpixels into GCNNs instead of individual pixels. This methodology enabled the training phase of GCNNs on an extensive assortment of pixels in HSIs, specifically addressing the challenges of land cover classification. However, it is noteworthy that these techniques need to fully address the inherent challenges associated with GCNNs. To bridge this gap, Hong et al. [225] proposed a novel Mini-GCN. As implied by its nomenclature, the Mini-GCN adopts a mini-batch training format akin to the methodology employed by CNNs. This innovation not only effectively reduces computational costs but also facilitates quantitative evaluation and integration with CNNs. The culmination of this research is the development of a fusion network (FuNet) for HSIC, representing a significant stride forward in the field.

E. Future Directions for CNN-Based HSIC

We reviewed the latest advancements in CNNs for HSIC in the previous sections. Even though CNN-based HSIC methodologies have accomplished significant progress in terms of classification results, there are still numerous considerations that require additional research. For example, there is a demand to continue working on these frameworks that can use spatial and spectral features for HSIC. Many of the methodologies discussed previously employ DR techniques to improve spectral–spatial representation, but these methods eliminate helpful HSI spectral features. As a result, stable HSIC methods capable of preserving spectral features are necessary. Although applying these methods enhances the computation complexity and slows down the training procedure, parallel processing of these networks utilizing field-programmable gate arrays (FPGAs) and graphical processing units (GPUs) represents a cutting-edge approach to optimize computational performance while maintaining model efficiency. By harnessing the parallel computing capabilities of FPGAs and GPUs, the network's operations can be distributed across multiple cores, allowing for the simultaneous execution of tasks and significantly reducing processing time.

Furthermore, as CNNs become deeper and deeper, extra labeled training data are needed for efficient classification, and as previously mentioned, HSI lacks labeled training data. To solve this problem, state-of-the-art research is needed to incorporate the CNN with unsupervised methods. In addition, we need to focus further on the generalization capabilities of CNNs, especially for the input data format. In order to create a more general CNN-based novel approach, GCNNs may be a suitable choice to integrate with CNNs. With the aid of this, we presume to be capable of further turning the efficiency bottleneck and producing HSIC that is more effective.

VIII. Autoencoders

AEs are an effective symmetrical neural network for HSIC because of their ability to unsupervised learning. The AE provides a compact feature representation of high-dimensional HSI data rather than performing a classification process on its own. As depicted in Fig. 6, the AE is composed of an input layer, a hidden or encoding layer, a decoding layer, and an output layer. When trained on input data, the AE can reconstruct the input by encoding it into a latent representation. The goal of AE is to decrease the reconstruction anomaly that results in the variation between the input and the output in order to obtain a compressed feature representation of the input data.

The SAE is formed by loading several AE layers so that the output of one layer serves as the input for the layer above it. A variation of AE called denoising autoencoder (DAE) shares the same structure as the AE with the exception of the input data. The output of DAE is the actual input signal without noise, whereas the input is distorted by noise addition. As a result,

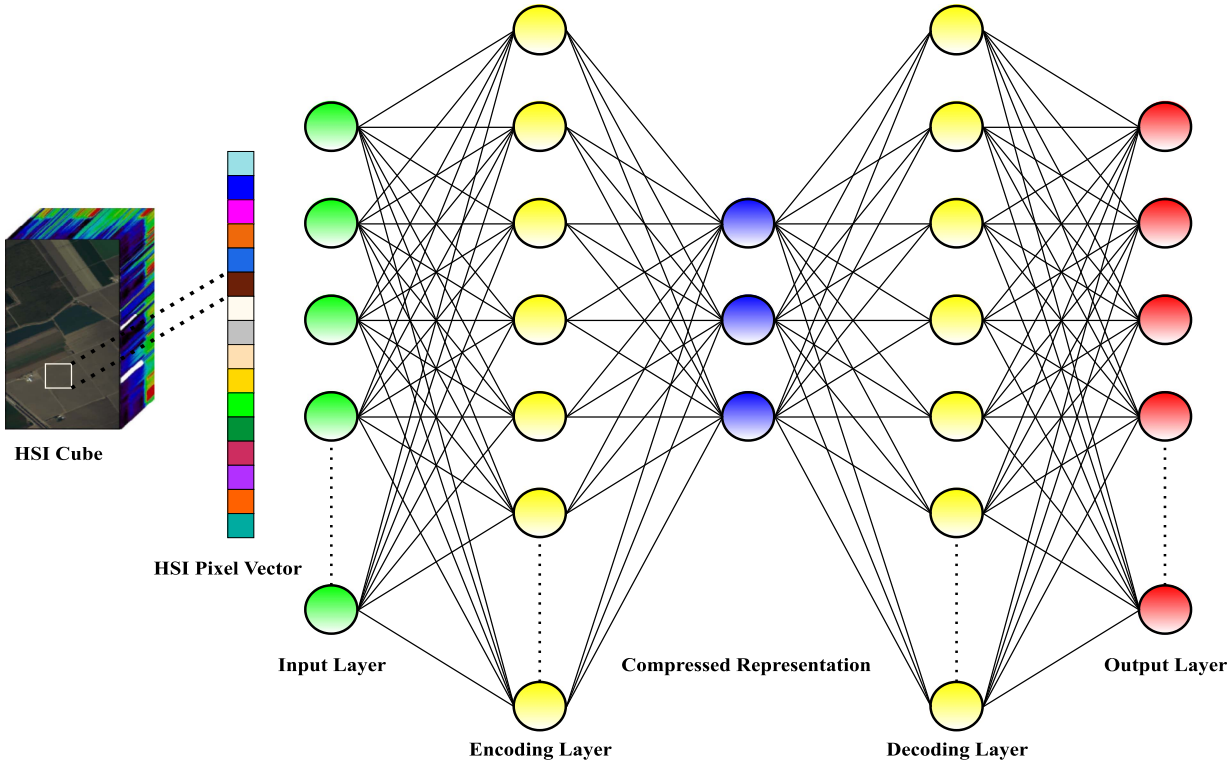


Fig. 6. Generic architecture of an autoencoder.

unlike the AE, the DAE can restore the actual input from a noisy input signal.

Zhu et al. [228] presented a fusion of multilayer AEs with higher noise fraction to extract high-level representation from the dataset, which lowers the spectral dimensions of HSI, whereas a SoftMax logistic regression classifier is used for HSIC. The research indicated in [229] integrated multimanifold learning [230] approach with a counteractive AE [231] for enhanced unsupervised HSIC. Zhang et al. [232] utilized spectral–spatial information of HSI using an unsupervised feature-extracting approach comprised of a recursive AE network. It retrieves features from the target pixel's neighborhood and assigns weights according to the spectral correlation between the target and neighboring pixels. Hao et al. [233] presented a two-stream DNN with a class-specific fusion framework that learns the fusion weights dynamically. A sequence of stacked DAEs is employed in the first stream to retrieve spectral information, and the CNN is applied in the second stream to retrieve spatial information. After completing the separate classifications from both streams, the final classification decision is reached through a process of merging the class prediction scores obtained from each stream's outcomes. This fusion of class predictions leverages the comprehensive information collected from multiple sources, enabling the model to make a more informed and accurate decision.

A further study introduced a hybrid approach for multifeature-based spectral–spatial HSIC that makes use of a sparse AE for high-level FE, guided filters [234] for spatial feature retrieval, and PCA for DR. The approach suggested in [235] adopted batch-based training of AEs, and features are produced by merging spectral and spatial features by a mean pooling mechanism. This combination of feature representations is a powerful approach that allows the model to capture both local and global information in the data. In a study [236], the researchers introduced a novel spectral–spatial HSIC approach, which involved extracting the appropriate spatial resolution from the HSI. To achieve this, they employed stacked sparse AEs for high-level FE to capture essential spectral–spatial information.

In a similar manner, Lv et al. [237] employed a stacked sparse AE for spectral–spatial and multifractal representations, as well as other higher order statistical representations. In [238], an integration of SAE and ELM is presented for HSIC, which partitions the training datasets and infuses them using SAE; after transformation, feature subsets are reorganized following the actual sequence of the training dataset and transmitted to ELM-based classifiers, with Q-statistics chosen for final classification performance. This feature subset computation enhances the reduction of variability among base classifiers. Correspondingly, Ahmad et al. [239] incorporated a robust and efficient multilayer ELM-based AE that acquires features in three folds, as presented for HSIC in [40].

To address the problem of increased intraclass variance and increased interclass correlation in HSI, Zhou et al. [240] created an SAE-based HSIC that can extract compressed and discriminatory information by introducing a local Fisher discriminant regularization. Consequently, the most recent findings [241] combine a k-sparse denoising AE with spectral-restricted spatial features to achieve the high intraclass variance of spatial features for HSIC. Paul and Kumar [242] introduced an HSIC framework that initially generates spectral segments of HSI based on a joint

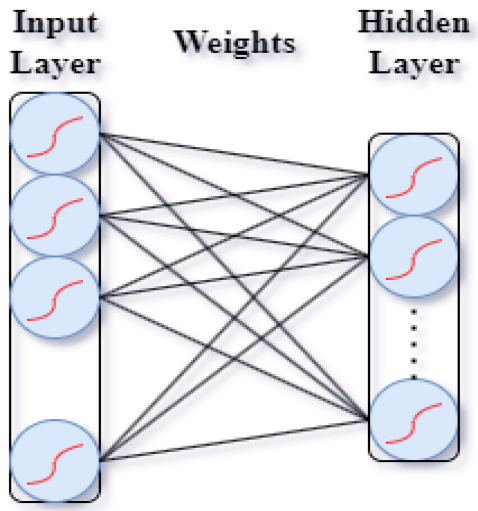


Fig. 7. Basic structure of a restricted Boltzmann machine.

feature criterion to decrease computational complexity during FE via SAE, then incorporates spatial features employing EMPs, and finally uses SVM/RF for classification purposes. Lately, Liu et al. [243] have employed an SAE to classify an oil slick on the sea surface by combining spectral–spatial HSI features.

A. Future Directions for AE-Based HSIC

In the previous sections, we reviewed the latest progress in AE-based HSIC methods. Though those methodologies exhibit strong predictive efficiency and generalization abilities, the more complex task is mainly required. Most of those described methods do not fully leverage immense spatial features, so additional methods that can leverage joint spatial and spectral features for HSIC must be established. Furthermore, high intraclass variance and high interclass correlation in HSI constrain the generalization of the classifier. Most of the research discussed in the previous section has tackled this problem, but more research is needed to address it. Additional research into methods such as pretraining, cotraining, and adaptive neural networks, among others, for AE-based HSIC frameworks could be one key area.

IX. Deep Belief Network

A DBN [244] represents a powerful type of hierarchical DNN that learns features in a layer-by-layer manner, progressively uncovering intricate representations of the input data. As illustrated in Fig. 7, the layers in DBN are constructed utilizing a restricted Boltzmann machine (RBM) with a two-layer architectural style, during which viewable units are associated with hidden units [237]. Zhang et al. [245] provided a thorough insight into the RBM. To retrieve fully detailed information from input data, the hidden unit of one RBM can be transmitted to the visible units of another RBM. The layer-by-layer structure employed in the DBN facilitates a recursive training process that empowers the network to acquire increasingly complex and abstract deep features from HSI. Fig. 8 depicts the structure of a three-layer DBN, showcasing the hierarchical arrangement of its layers.

In the literature, a number of research studies applied the DBN for HSIC. For example, Ayhan and Kwan [246] combined spectral–spatial features with the DBN for land cover classification and compared their results to the state-of-the-art classification methods. The DBN typically learns in two stages, the first of which is unsupervised pretraining using unlabeled instances and the second is supervised fine-tuning using labeled instances. However, such a training procedure could lead to two issues: 1) because of coadaptation [247], multiple hidden units may have a tendency to behave in similar fashion [248]; 2) because of the sparseness and specificity of activated neurons, a few of which may always be dead or interacting [249]. In order to address these two issues, Zhong et al. [250] regularized the pretraining and fine-tuning operation by enforcing a diversification before improving the DBN's classification performance for HSI.

In [251], an innovative DBN-based texture feature optimization method was proposed, aiming to elevate the performance of texture analysis tasks. This approach goes beyond conventional methods by incorporating band grouping and sample band selection techniques seamlessly combined with the application of a guided filter. A DBN model then learns the enhanced texture features, and the final classification outcomes are generated by a SoftMax classifier. The study in [252] incorporated a parallel layer approach with a logistic regression layer for classification and a Gaussian–Bernoulli RBM to retrieve high, local invariant, and nonlinear features from HSI.

Several research studies are regarded to jointly incorporate the spectral and spatial features present in HSI in order to enhance classification performance. For example, Li et al. [253] proposed a DBN approach with the logistics regression layer and presented that the joint exploitation of spectral information increases classification performance. Consequently, Sellami and Farah [254] presented a spectral–spatial graph-based RBM technique for HSIC, which builds the spectral–spatial graph by measuring combined correlation premised on spectral and spatial information; then, the model proceeds with training an RBM to extract meaningful joint spectral–spatial features from HSI. Subsequently, these extracted features are fed into a DBN alongside a logistic regression layer, enabling the final classification step.

A. Future Directions for DBN-Based HSIC

In the previous section, we surveyed the most recent advancements in DBN-based HSIC approaches. We noticed that in comparison to numerous different DNNs, only several studies had used DBNs for HSIC. Consequently, it is essential to consider further DBN-based rigorous methods for HSIC that can utilize both spatial and spectral features. One avenue of exploration involves the integration of DBNs with cutting-edge architectures, such as knowledge distillation and transformer models, aiming to leverage the strengths of these architectures for improved classification accuracy. Researchers may delve into unsupervised and semisupervised learning scenarios, investigating how DBNs can effectively utilize unlabeled or partially labeled hyperspectral data.

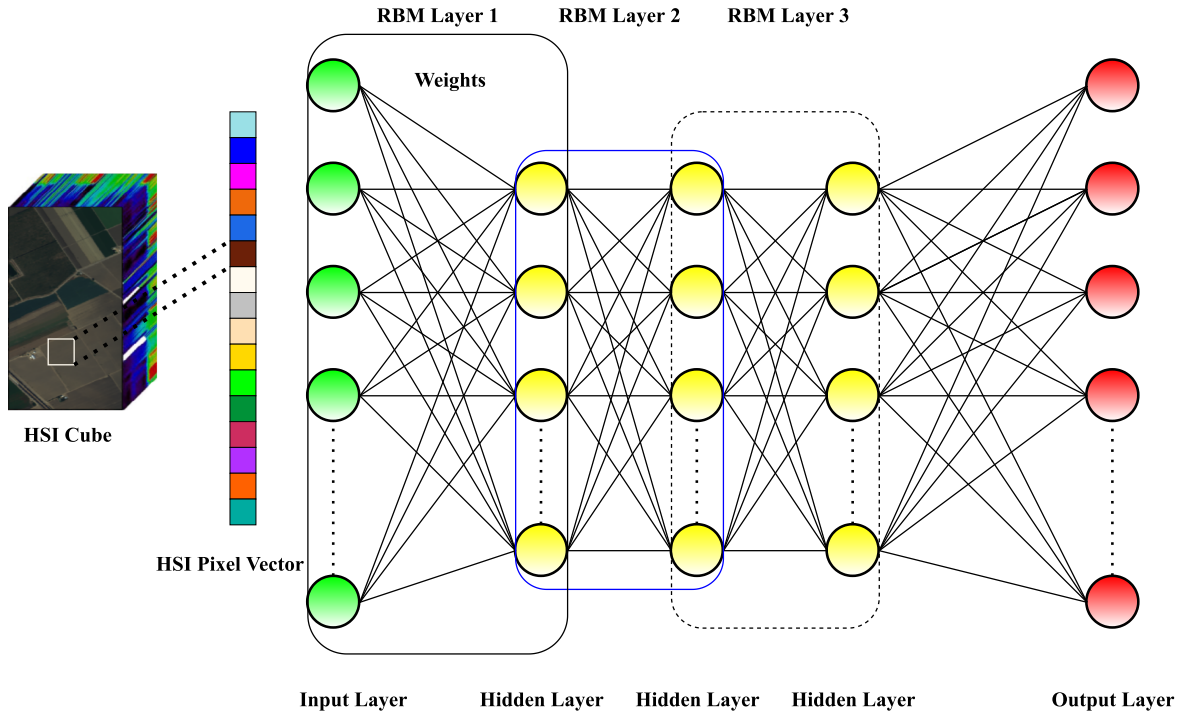


Fig. 8. Three-layer DBN architecture.

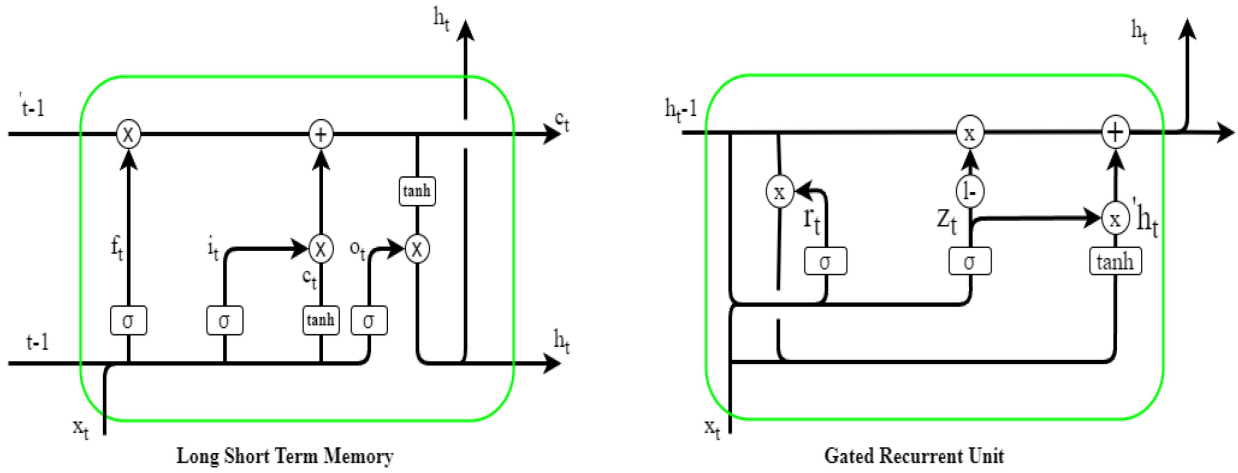


Fig. 9. Internal structure of long short-term memory and gated recurrent unit.

In addition, adapting DBNs to perform robustly with limited datasets is crucial, prompting the exploration of regularization techniques and transfer learning (TL) strategies tailored for the unique challenges posed by hyperspectral images. The incorporation of spatial information into DBNs and the exploration of attention mechanisms can enhance their ability to capture both spectral and spatial features, contributing to more comprehensive classification results. Further research may focus on domain adaptation, allowing DBNs to generalize across different hyperspectral datasets and the integration of multimodal data to harness complementary information from diverse sources. Another suggestion for future research could be the regularization of DBN pretraining and fine-tuning mechanisms to effectively solve the problem of dead neurons.

X. Recurrent Neural Network

The structure of RNNs, depicted in Fig. 9, consists of loop connections in which the node activation of the subsequent phase depends on the preceding phase [255]. Thus, RNNs have the capability to learn temporal sequences. RNN architectures operate spectral features from HSI data as a time sequence, with spectral bands acting as time strides [256]. RNNs are classified into three types: 1) Vanilla; 2) long short-term memory (LSTM); and 3) gated recurrent unit (GRU).

Vanilla is the most basic RNN framework and is prone to results in feature loss when handling high-dimensional data. LSTM models with two phases solve such problems by managing the data stream via three gates: input, forget, and output.

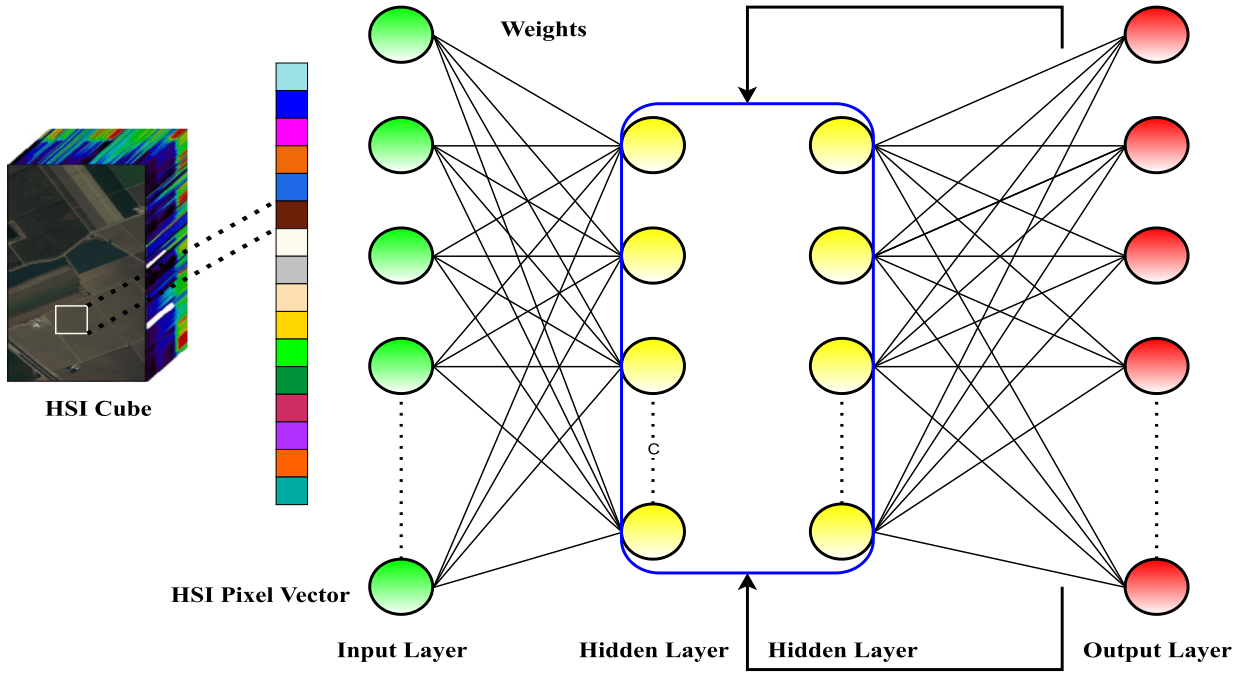


Fig. 10. Recurrent neural network.

It learns the related data by substituting the irrelevant data over a period. While the gate-controlling approach provides the LSTM with a notably dynamic strategy, the GRU version maintains the consistency of the Vanilla model and, at the same time, achieves enhanced accuracy compared to the LSTM. GRU is a simplified variant of LSTM that replaces the output gate with an update (zt) and reset (rt) gate and modulates the input and forget gate. Fig. 10 shows a description of the internal architectures of LSTM and GRU.

Hang et al. [72] introduced a novel RNN-based methodology for HSIC. This approach incorporates a unique activation function called "parametric rectified tanh" and integrates GRUs. The model effectively utilizes the sequential features present in the HSI data to specify class labels, enhancing its capability to capture complex patterns and relationships within the hyperspectral data. A new LSS technique following the RNN approach was presented in [145], which initially retrieves low-level information from HSI employing the GF and DMPs [146] and further merges this information to acquire LSS features from the presented technique; these LSS features are then transmitted to an RNN architecture to retrieve high-level features, with a SoftMax layer employed for classification purpose.

Considering the versatility of spatial features in improving classification accuracy, Zhou et al. [257] introduced a spectral–spatial LSTM-based network that extracts spectral and spatial features of HSI by leveraging two distinct LSTM preceded SoftMax layers for classification, with an ensemble learning approach applied to obtain joint spectral–spatial classification performance. Besides that, Sharma et al. [258] introduced a patch-based RNN with LSTM cells for land cover classification that incorporates multitemporal and multispectral features as well as spatial features.

In literature, many studies have presented CNN-based hybrid RNN (CRNN) architectures for HSIC. For example, Wu and Prasad [180] applied a convolutional recurrent neural network (RNN-CNN) architecture in their approach. In their methodology, the initial convolutional layers are utilized to extract position-invariant middle-level features, while the subsequent recurrent layers are favored for retrieving spectral-contextual information in HSIC. Similarly, Wu and Prasad [259] used pseudolabels to implement the related model for semisupervised HSIC. Zhou et al. [260] proposed an HSIC approach in which the CNN is employed to retrieve spatial information from HSI, and this information is afterward sent to a GRU-based FuNet that conducts feature-level and decision-level fusion.

In a similar way, Luo et al. [261] leveraged both the spatial and spectral data inherent in HSI by integrating a CNN with a parallel GRU-based RNN. This innovative amalgamation serves to streamline the training of the GRU and enhances the overall efficiency of the process. In another study, a bidirectional convolutional long short-term memory (CLSTM) model was proposed in [155] with the express purpose of jointly harnessing the spectral–spatial feature of HSI for the process of classification. Luo [261] combined multiscale local spectral–spatial features obtained through a 3-D CNN with a hierarchical RNN. This hierarchical RNN was designed to scrutinize the spatial correlations of local spectral–spatial features at multiple scales.

The introduction of both recurrent 2-D CNN and recurrent 3-D CNN for HSIC in [262] proved noteworthy, given the intriguing correlation observed between these methodologies and the respective 2-D and 3-D CNN models. This correlation affirms the significance of recurrent CNN in the context of HSIC. In the study [263], the CNN and CLSTM were synergistically combined. First, the authors harnessed the capabilities of a 3-D CNN to capture essential low-level spectral–spatial features

from the raw hyperspectral data. Subsequently, the researchers employed a CLSTM network, a specialized variant of the RNN, to repeatedly analyze the obtained low-level spectral–spatial features. The CLSTM is uniquely suited to process sequential data while retaining the ability to capture long-term dependencies. In a recent study by Hang et al. [72], a cascade RNN approach was proposed for HSIC. This novel architecture comprises two layers of GRU-based RNN. The first layer focuses on minimizing redundant spectral bands, aiming to enhance the efficiency of the subsequent FE process. On the other hand, the second layer is dedicated to acquiring informative features from the HSI, ultimately facilitating accurate classification. In addition, several convolutional layers were incorporated to assimilate the rich spatial details stored within the HSI. These advancements reflect the ongoing evolution and innovation in the field of HSIC.

A. Future Directions for RNN-Based HSIC

In the previous section, we reviewed the latest progress in AE-based HSIC methods. However, RNN-based HSIC architectures have received much interest from the remote sensing society and have shown excellent classification results; several attributes still require more research, such as the development of sequential input data for RNNs. The majority of the reviewed techniques focused on the HSI pixel, treating it as a sequential point representing a data sequence formed by the pixel values from each spectral band. However, this substantially lengthens the input sequence for RNNs, which can cause an overfitting problem.

In addition, processing these lengthy data sequences tends to grow computational time and slows learning. In order to acquire better classification performance of RNN-based HSIC, the implementation of parallel processing techniques requires to be thoroughly explored. Furthermore, methodologies such as clustering spectral bands to reduce data sequence distance and utilizing a whole spectral signature to improve discrimination between different classes can be investigated more to build the RNN model's sequential input. Another promising future direction may be the integration of RNN-based HSIC architectures in a real-world multitemporal HSI context, as this could yield significant advancements in various applications. By incorporating RNNs with HSIC in the analysis of multitemporal hyperspectral imagery, researchers and practitioners could unlock a plethora of new possibilities.

XI. Approaches for Addressing Limited-Labeled Data

However, despite the fact that DNNs have been effectively used for the HSIC task, they need a sizable portion of labeled training data. Furthermore, as previously mentioned, the selection of labeled HSI is crucial and expensive, involving various factors that either necessitate the expertise of human professionals or the investigation of real-time situations. The scarcity of labeled training data impedes classification results. Numerous innovative approaches have been presented in the literature to address the aforementioned problem. In this section, we will comprehensively explain several such approaches and techniques while aiming at active learning (AL) algorithms and ensemble-based learning methodologies.

A. Data Augmentation

DA has been shown to be an efficient method for HSIC in dealing with the problem of insufficient training samples. It produces novel samples from the actual training samples while incurring no extra labeling costs. DA techniques are divided into two categories: 1) data wrapping and 2) oversampling [264]. Data wrapping typically encodes multiple invariances (translational, size, and illumination) while sustaining the labels, whereas oversampling-based augmentation techniques increase the training dataset by producing artificial sample sizes premised on actual data distributions. Methods for oversampling involve mixture-based instance generation [264], feature space augmentations [265], and generative adversarial networks (GANs).

According to the HSIC literary works, multiple DA-based approaches have been used to enhance classification efficiency by mitigating the overfitting problem, which commonly occurs due to a lack of training data. For example, Yu et al. [266] augmented the training data with three DA operations (flip, rotate, and translation) and then used this augmented data to train CNN for HSIC. Li et al. [267] introduced a thorough comparative study of numerous widely used HSI DA methods and presented a pixel-block pair-based DA that used both spectral and spatial HIS features to produce unique examples and train a CNN model for HSIC. Cao et al. [268] assessed the classification efficiency of a CNN in combination with AL both with and without DA techniques. Their findings revealed that the inclusion of DA led to improved classification accuracies, demonstrating its significant contribution to enhancing the model's performance. Consequently, in the study [269], the DA-based CNN outperformed a PCA-based CNN model by 10% in HSIC accuracy.

The mentioned techniques previously employed offline DA methods to boost training data by generating novel examples before a model's training phase. Lately, a new DA approach for HSI has been presented in [270], which produces the sample sizes at test time instead of artificially boosting the training data, and a DNN trained over actual training data, as well as a voting mechanism, has been utilized for the concluding class label. Nalepa et al. [270] also presented two accelerated DA methods for high-quality data syncretization to enhance the predictive performance of DNNs. Nalepa et al. [271] proposed a relevant PCA-based online DA approach that reconstructs novel examples during inference rather than training.

B. Semisupervised Learning

SSL techniques leverage both labeled and unlabeled data to learn the underlying data distribution. By integrating unlabeled and labeled instances during training, these methods expand the pool of training examples and establish meaningful correlations between the feature space and class labels. Numerous SSL approaches for HSIC have been proposed in the research, falling under several broad categories: cotraining, self-training, GANs, graph-based SSL models, and semisupervised SVMs are among the commonly employed methods. A latest thorough review of

these SSL methods can be explored in [272]. Furthermore, Pise and Kulkarni [273] also presented a comprehensive analysis of SSL methodologies.

There is a brief summary of the SSL-based HSIC methods in [274], where the authors also compare these methodologies in great depth. The approach described in [259] pretrained a CRNN for HSIC using pseudo or cluster-labeled instances and then fine-tuned the network using limited labeled data. Likewise, Kang et al. [160] presented a semi-HSIC approach that uses PCA and increased morphological attribute profiles to retrieve pseudolabeled sample sizes that are injected into a CNN-based DFFN.

Fang et al. [275] introduced a dual cotraining method focused on spectral and spatial features of HSI. Consequently, Zhou et al. [276] pretrained two SAEs individually, the first pretrained model utilizing spectral information and the second pretrained model utilizing spatial HSI information, and fine-tuning is obtained through a cotraining strategy. To improve the training data, Li et al. [277] introduced a region-information-based self-training strategy. In [278], an innovative approach to self-training was proposed, which hinges upon the principles of graph theory. This methodology commences with a preliminary sample size, which is ascertained through the application of subtractive clustering, a technique notable for its effectiveness in identifying cluster centers in high-dimensional data. Wu et al. [161] made considerable strides in enhancing the outcomes of HSIC by employing a novel strategy of pseudolabeling unlabeled datasets. This was accomplished using a framework rooted in the principles of clustering-based self-training. An additional layer of finesse was added by regulating the self-training process with spatial constraints, thereby ensuring the relevance and appropriateness of the pseudolabels in the spatial context.

C. Few-Shot Learning

Few-shot learning (FSL) is a machine learning paradigm that aims to train models to recognize and classify new classes with only a small number of examples per class [279]. This is particularly useful in scenarios where obtaining a large labeled dataset is challenging or expensive. HSIC involves categorizing each pixel in a hyperspectral image into predefined classes based on the spectral information. In HSIC, where the number of classes may be large and diverse, a model trained using FSL may generalize better to new classes as it learns to capture the underlying spectral patterns.

Several studies have introduced FSL in HSIC to boost performance in limited data conditions. For example, Liu et al. [280] proposed a novel deep FSL method to tackle the limited labeled samples in HSI classification. The method employs three key strategies: 1) using a deep residual 3-D CNN; 2) training the network through episodes to learn a metric space; and 3) using the nearest neighbor classifier within the learned metric space. In another study [393], the authors proposed a novel approach to heterogeneous FSL for HSIC, utilizing only a limited number of labeled samples per class . A few studies introduced a graph-based approach that has also been integrated into FSL [281], [282], [283], [284].

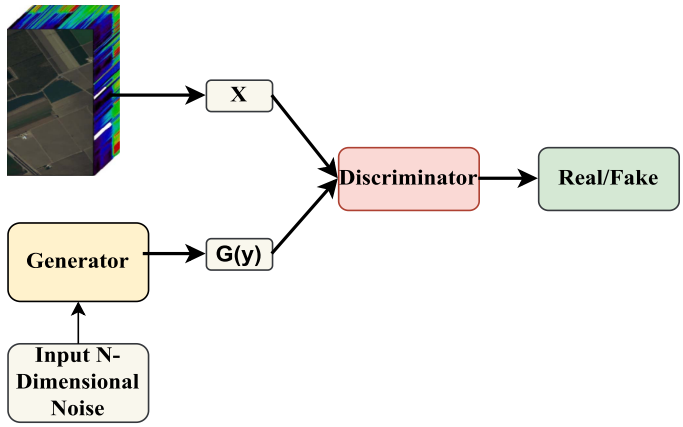


Fig. 11. Generic framework of a generative adversarial network.

Recently, state-of-the-art methods have been proposed with the integration of FSL for HSIC. For example, Ran et al. [285] proposed an innovative classification framework, i.e., deep transformer and FSL, with the aim of achieving fine-grained classification of HSI using only a limited number of instances in an FSL setting. Peng et al. [286] introduce a novel convolutional-transformer-based FSL method. It concurrently applies FSL in source and target domains, establishes a consistent scenario, aligns domain dimensions using a domain aligner, and employs a convolutional transformer network for local–global FE. Huang et al. [287] introduce HFC-SST, an advanced spatial–spectral transformer (SST) for few-shot HSIC. Effectively capturing local spatial–spectral information with limited labeled samples, it employs a sequence generation method based on spatial–spectral correlation analysis and adjacent position information. In addition, an FE network, based on the transformer, further exploits spatial–spectral features within the input sequence.

D. Generative Adversarial Network

The GAN presented by Creswell et al. [288] consists of two DNNs; the first part of GAN is referred to as a generator, and the second part is called a discriminator, as shown in Fig. 11. By utilizing the data distribution information, GANs can learn to replicate the sample data. A spectral-feature-based GAN for SSL-based HSIC was presented in [289].

Consequently, He et al. [290] presented a spectral–spatial HSIC approach based on the GAN. Zhu et al. [291] built CNN-based 1-D GAN and 3-D GAN frameworks to improve classification efficiency for HSIC. The spectral features are generated using a 1-D customized GAN [292], in which the CNN employs for FE, and then, a majority voting mechanism is applied to produce HSIC. Feng et al. [293] have lately presented a novel multiclass spatial–spectral GAN, which employs duplet generator models to generate spatial and spectral features with the assistance of several adversarial strategies. Zhong et al. [294] introduced a novel conditional random field generative adversarial network framework that integrates the GAN model with conditional random fields to solve the data sparsity issue for HSIC.

Correspondingly, Wang et al. [295] explored a caps-triple-GAN model, which efficiently produces novel sample sizes utilizing a 1-D architecture triple GAN (Triple-GAN) and classifies the produced HSI samples employing capsule network (CapsNet). Xue [296] introduced using a novel 3-D-CNN-based generator network and a 3-D deep-residual-network-based discriminator network for HSIC. In [297], an integration of CapsNet and convolutional-LSTM-based discriminator model was presented for HSIC to extract high-level context information.

The study [298] addresses the insufficiency of training instances by employing a GAN model in which the discriminator's efficiency is enhanced more by an additional classification model to generate more architecturally cohesive virtual training instances. In addition, Roy et al. [299] introduced a generative-adversarial-minority-oversampling-based method to tackle the lasting issue of classwise data scarcity applied by HSIC to improve predictive accuracy. Hong et al. [300] proposed HighDAN, a high-resolution domain adaptation network, to improve the AI model's generalization across various urban environments. HighDAN preserves urban scene structure through high-to-low resolution fusion, addressing differences in remote sensing images between cities via adversarial learning.

E. Transfer Learning

TL increases a model's productivity and operational efficiency with previous insights of a related main goal to conduct a secondary task. Moreover, features retrieved from the appropriate source domain are transferred to the desired domain in order to learn previously unseen information. As a result, TL can be used successfully in applications with little or insufficient training datasets. TL approaches are further classified as supervised or unsupervised depending on the accessibility of labeled training examples. In general, both the source and target domains are supposed to be relevant but not identical. Although they have high variability, like, in the particular instance of HSIC, where the classes of concern are identical, data in the two domains may differ given the various acquiring conditions.

With DNN-based HSIC, the model extracts features in a hierarchy fashion, with bottom layers typically extracting generalized features when trained on a variety of examples. As a result, a novel classifier can be learned for the intended dataset using the features these layers have already learned. For example, Yang et al. [301] adapted the bottom layers of the pretrained model to the intended network for the accurate classification performance of the target HSI. They pretrained a two-branch spectral–spatial CNN architecture with a sufficient number of training examples from other HSIs. The intended network's upper layers are arbitrarily initialized in order to acquire the task-based information, and the network in its entirety is tweaked using a small number of labeled examples of the target HSI. In a similar fashion, Windrim et al. [302] introduced a proper technique for pretraining and fine-tuning a CNN before using it to classify novel HSIs. The research work in [303] merged DA and TL techniques to enhance HSIC efficiency in the context of a lack of training examples.

As previously described, data in the source and target domains can also vary in several ways; for example, in the situation of HSIs, the dimensions of two HSIs may differ because they were acquired from various sensors. The process of dealing with such cross-domain variability and transformation of information between them is referred to as heterogeneous TL (a comprehensive review of similar techniques can be explored in [304]). Many studies have been presented in HSIC research to solve the problem of transferring information between two HSIs with highly variable dimensions and distributions.

For instance, Lin et al. [305] presented an improved heterogeneous TL-based HSIC approach that performs effectively with both homogeneous and heterogeneous HSIs. In a research study, Li et al. [306] introduced a novel heterogeneous TL with a recursive reweighting framework for HSIC. In addition, Liu and Xiao [307] presented a band-selection-based TL method to pretrain a CNN that keeps a relatively similar amount of dimensions for different HSIs. Moreover, Lin et al. [308] introduced an unsupervised TL method to classify entirely unexplored target HSI, and de Lima and Marfurt [309] showed that networks trained on natural images can outperform networks trained from scratch utilizing relatively small HSI dataset for remote sensing data classification. For the first time, Hong et al. [310] introduced SpectralGPT, a pioneering universal remote sensing foundation model specifically designed to process spectral remote sensing images through an innovative 3-D generative pretrained transformer.

F. Transformer Models

Transformer models have demonstrated significant success in various natural language processing tasks, but their application has extended beyond text to computer vision tasks, including HSIC. The architecture of transformer models, originally designed for sequence-to-sequence tasks, has proven effective in capturing spatial and spectral dependencies in HSIC. Several studies demonstrated effective results of using transformer architecture in HSIC. For example, He et al. [311] introduced a novel classification framework named SST for hyperspectral image classification. The SST employs a carefully crafted CNN to extract spatial features, incorporates a modified transformer known as DenseTransformer to capture sequential spectra relationships with dense connections, and utilizes a multilayer perceptron to perform the final classification task. Qing et al. [312] presented SAT Net, an end-to-end transformer model designed for HSIC, leveraging the self-attention mechanism. This model incorporates both spectral attention and self-attention mechanisms to extract spectral–spatial features from the HSI image. Similarly, the authors of [313], [314], and [315] employed spectral–spatial transformer models to extract spectral and spatial features for HSIC.

Recently, many state-of-the-art models have been proposed with the integration of transformer models. For example, Yang et al. [316] introduce FusionNet, a hybrid network that integrates convolution and transformer architectures for HSIC. This fusion occurs through both serial and parallel mechanisms, maximizing the utilization of HSI features. Similarly, Zhang et al. [317]

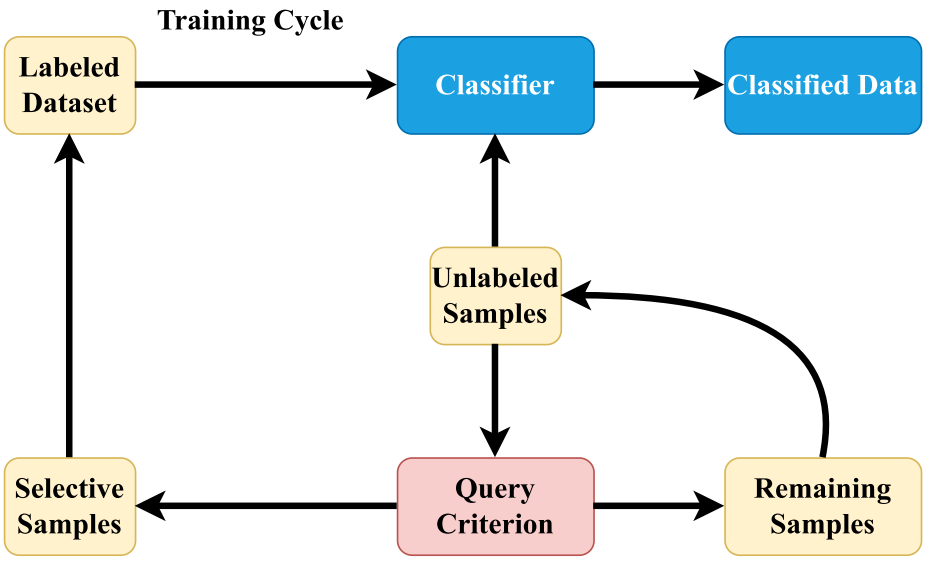


Fig. 12. Generic overview of active learning.

introduce the convolution transformer mixer (CTMixer), a comprehensive framework proposed for HSIC. CTMixer aims to seamlessly integrate the benefits of both convolutional and transformer paradigms for effective classification. Zhao et al. [318] introduce a novel framework called multiattention transformer with adaptive superpixel segmentation-based active learning. This framework demonstrates outstanding classification performance, particularly in scenarios with limited sample sizes. Qiu et al. [319] present a cross-channel dynamic spatial–spectral fusion transformer (CDSFT). Within the CDSFT, multiscale and multichannel features are initially extracted, followed by the derivation of cross-channel global features through transpose multihead self-attention.

Moreover, state-of-the-art variants of transformer models have been proposed to enhance the classification efficiency further and perform well in limited samples situations. For example, Roy et al. [320], introduce morphFormer, a novel morphological transformer. The model incorporates a trainable spectral and spatial morphological network, employing morphological convolution operations alongside the attention mechanism. Hao et al. [321] introduce a transformer with residual upscale (RU) GAN, consisting of a generator (G) and a discriminator (D). Within the generator (G), a technique called RU is suggested to enhance the resolution of generated features. In addition, RU facilitates the extraction of texture features and the capture of contextual relationships. Moreover, Cao et al. [322] introduce a transformer-based model using contrastive learning for enhanced performance. The model aims to integrate both methods and enhance overall performance. Liang et al. [323] propose a swift HSIC method that combines transformers and SimAM-based CNNs. The SimAM-based CNN utilizes an improved hierarchical 2-D dense network structure to capture intricate spatial characteristics. A dual attention unit directs the model's focus to discriminative spatial pixel characteristics and efficient feature map channels, suppressing irrelevant information for classification.

G. Active Learning

AL improves a classifier's prediction accuracy in an iterative manner by effectively enhancing the capacity of the training dataset per training epoch process using an unlabeled pool of instances. For every epoch, AL improves the training dataset by effectively choosing the more beneficial examples from the pool of unlabeled data, and an oracle (human- or machine-based) allocates the actual class labels to such examples. Subsequently, the helpful examples are included in the training dataset, and the model is retrained while using the novel training dataset. The procedure is iterated unless a termination criterion is met, which could be the sample of the training dataset, the total number of iterations, or the required classification result. Fig. 12 depicts a basic framework of AL.

To boost performance, the valuable examples set are chosen in such a manner that they are helpful and reflective of the general input distribution. AL approaches are classified as stream-based or pool-based based on the considerations for introducing novel examples to the training dataset. In stream-based choice, one example is taken at a period from a true set of unlabeled examples, and the method determines whether or not to label it depending on its applicability. In a pool-based approach, examples are queried from a pool of unlabeled datasets depending on scoring results calculated from different metrics to assess the effectiveness of the examples.

Ganti and Gray [324] discovered that streamed-based choice yields lower learning rates than pool-based choice because the previous approach leads to more query examples. In pool-based choice, it is critical to include diversification in the pool of instances to prevent redundant data within the pool of instances. In general, three factors are considered when choosing the most useful instances: heterogeneity behavior, prediction accuracy, and instance representativeness. A detailed overview of such instances techniques is given as follows.

1) *Heterogeneity-based methods:* These techniques choose instances that are more diverse than previously described

instances in terms of model diversification, classification unpredictability, and contention among a group of different classifiers. Instances of heterogeneity-based frameworks include uncertainty sampling, intended model alteration, and query by committee.

a) *Uncertainty sampling:* In this framework, the classification model progressively seeks to query labels for instances where uncertainty is highest during predictions. Samples are chosen based on scores exceeding a predefined threshold, and those with the highest ratings are queried for labels. A straightforward approach involves applying a probabilistic classifier querying the label if the correctly classified probability is close to 0.5 in a binary classification case.

b) *Query by committee:* These heterogeneity-based frameworks focus on the sampling procedure on differences in the predictions of different classification models via the relatively close set of labeled instances. To predict the class labels of unlabeled instances, a committee of different classification models trained on a similar set of training data sizes is applied, and the instances for which classification models vary the most are chosen for querying labels. The committee of various classification models can be constructed utilizing ensemble learning approaches such as stacking and Boosting [325] or switching the hyperparameters [326]. In general, a smaller number of diversified classification models is sufficient for forming a committee [325], [327].

c) *Model variability:* A heterogeneity-based framework selects examples that differ significantly from the existing model in the context of the gradient of the objective function. These methods tend to query the label for examples that vary significantly from the existing model. Such sampling methods are solely applicable to models that use gradient-based training and optimization processes.

2) *Performance-based methods:* These techniques take into account the impact of including queried examples on prediction accuracy. They attempt to improve prediction accuracy by lowering variance and error. Performance-based sampling is divided into two categories.

a) *Error minimization:* This method, linked to uncertainty sampling, employs metrics that increase the label probability for querying examples, while a decrease in predicted error rate reduces label unpredictability. In the context of binary classification, minimizing error rates involves selecting examples with probabilities exceeding 0.5. These approaches are often referred to as the highest certainty models [326].

b) *Variance minimization:* This method, linked to uncertainty sampling, boosts label probability for querying examples based on unpredictability metrics. In binary classification, minimizing error rates selects examples with probabilities greater than 0.5—referred to as the greatest certainty models [328].

3) *Representatives-based methods:* Heterogeneity-based methods may include anomalies, while performance-based techniques, like accurate error prediction, naturally mitigate such instances. Representative sampling, focusing on common examples in the input distribution, avoids outliers. Density-weighted methods, such as information density, consider both heterogeneity and representativeness in capturing dense input regions during query processing [326].

Recently, AL has significantly empowered HSIC. Liu et al. [329] introduced the feature-driven active learning method, tailored for HSIC, while Zhang et al. [330] proposed active semisupervised RF, leveraging spectral–spatial features. These techniques dynamically select valuable examples for the training set, showcasing the potent integration of AL with HSIC. Several AL-based HSIC approaches have incorporated spatial features. Guo et al. [331] combined spectral and spatial properties of superpixels, Xue et al. [332] used neighborhood and superpixel features to increase data uncertainty, and Bhardwaj et al. [333] integrated spatial features using attribute profile information in an AL-based HSIC approach.

Batch-mode AL approaches are extensively used to speed up the learning process. In every iterative process, these frameworks choose a batch of examples to be queried for a label. As a result, in batch mode AL methods, sample diversification is essential in order to prevent redundant data. Patra et al. [334] presented a multicriteria batch-mode AL technique that describes a novel query function based on diversification, unpredictability, and cluster assumption measures. Such factors are described by utilizing the attributes of KNN, SVM, and K-means clustering, and genetic algorithms are then employed to select the batch of highly impactful examples. Correspondingly, Zhang and Crawford [335] introduced a regularized multimetric batch-mode AL approach for HSIC that takes advantage of several HSI features.

Xu et al. [336] introduced a multiview AL (MVAL) approach that evaluates the instance from diverse perspectives and estimates the information quality of the example using multiview intensity-based query specifications. In a similar manner, Pradhan et al. [337] used the Fisher discriminant ratio to produce various views, utilizing the approach of multilearning. In another research study, Zhang et al. [338] introduced a new method of adaptive MVAL approach for HSIC that leverages both spatial and spectral information in each view. Lately, Li et al. [339] have presented a new SPSMVAL method that produces different viewpoints for HSIC using pixel-level, subpixel-level, and superpixel-level information. In addition, the presented technique uses joint posterior probability estimation and differences between various viewpoints to query the representative examples.

Recently, a number of studies have integrated the AL and DNN frameworks. For example, Liu et al. [340] introduced a novel ADL framework for HSIC, and Sun et al. [341] combined AE with AL methods. In addition, Haut et al. [342] proposed a novel B-ALCNN method by combining Bayesian CNN with the AL framework for spectral–spatial HSIC. Moreover, Cao et al. [268] introduced a novel ALCNN approach to effectively leverage unlabeled examples for HSIC. Several studies combined AL

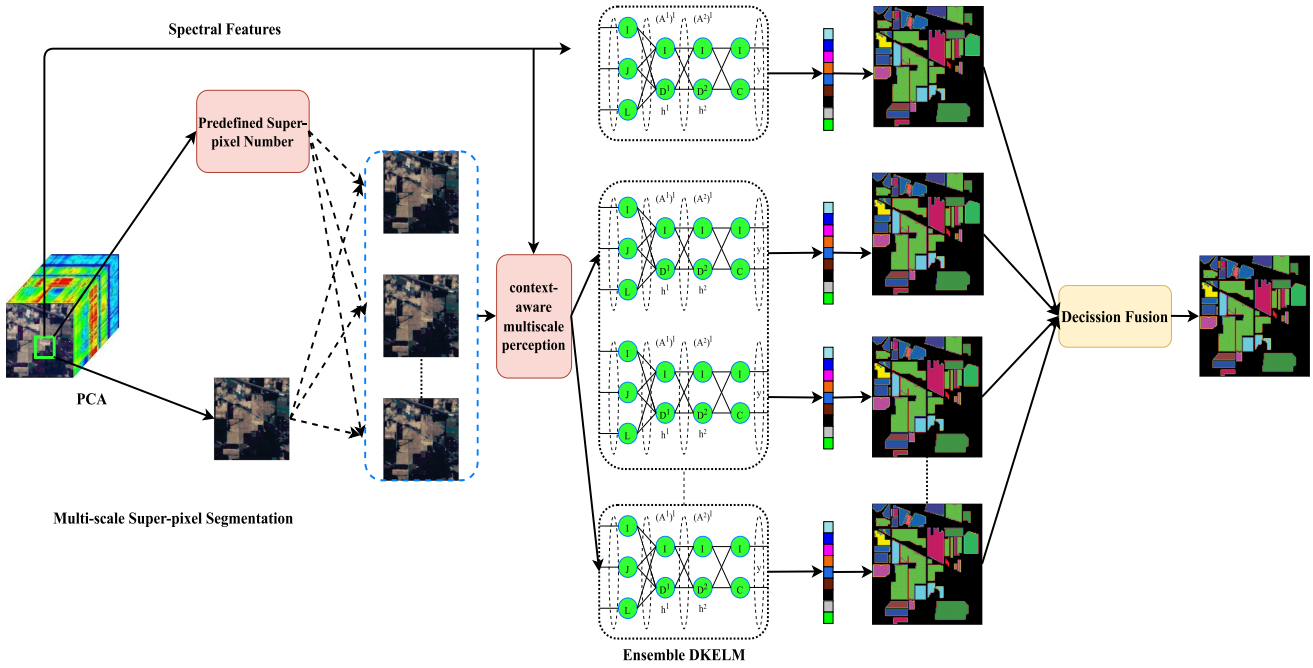


Fig. 13. Deep ensemble learning framework.

and TL for HSIC. In order to accurately reflect the data from the source and target domains, Lin et al.'s [343] AL-TL approach retrieves the salient examples and makes use of high-level features. In a different study [344], a stacked-sparse-AE-based active TL method was put forth that makes use of both spectral and spatial features for HSIC. Another study [345] proposed an active multikernel domain adaptation that merged domain adaptation and AL techniques for HSIC based on several kernels.

AL-based HSIC advances model generalization. For example, Ahmad et al. [36] introduced fuzziness-based AL (FAL) to enhance the generalization of discriminative and generative models. FAL calculates fuzziness-based depth, selecting training set representatives with high fuzziness scores and small distances from class boundaries. Ahmad et al. [346] introduced MNSAL for multiclass HSI classification, combining spatial prior fuzziness with MLR through partition and augmented Lagrangian methods. They conducted a comprehensive comparative analysis with state-of-the-art sample choice techniques and classifiers. Yao et al. [347] suggested a semisupervised CNN framework named SA-CNNs, utilizing AL and superpixel segmentation for HSI classification. They initiate the process by pretraining a CNN model on a limited unbiased labeled dataset, followed by inferring unlabeled data to generate pseudolabels.

H. Ensemble and Deep Ensemble Learning

Ensemble learning integrates some "weak" classification models using the appropriate approaches to produce better than any classification model. Boosting [348], Adaboost [349], and RF [350] are a few ensemble approaches that have been presented. The ensemble classification, which signifies a mono hypothesis, is versatile in terms of the characteristics it can depict. In theory and practical application, ensemble approaches generally enhance classification accuracy. Fig. 13 shows the deep ensemble framework.

Numerous ensemble-based approaches for HSI classification have been introduced in regard to hyperspectral remote sensing. For instance, to investigate the capabilities of ensemble learning for HSI classification, Waske et al. [351] presented a multiple classifier framework that integrates an SVM and random feature selection. Yao et al. [352] proposed a lightweight ensemble model with the integration of spectral features' refining module. The refining module approach is based on the method of attention approach. Moreover, in another study, Nalepa et al. [353] proposed deep ensemble with the integration of different CNN architectures. Furthermore, they also introduce model augmentation methods that enable the synthesis of novel DNNs based on the actual model by integrating the Gaussian noise into the model weights.

In [354], a novel deep ensemble CNN has been presented with the extension of the samples of HSI classifications. The spatial features are first retrieved and fused with the spectral features in order to enable the classifiers to acquire spatial and spectral features. They also utilized the PPF approach to increase the total size of training samples. Chen et al. [355] proposed a deep ensemble architecture that consists of a DL model and subspace-based ensemble learning. Especially, two DL ensemble-based methods have been used to classify the HSI (CNN ensemble and deep residual ensemble). CNNs and deep residuals have been used as separate classifiers, and random subspaces have been employed to diversify the ensemble mechanism in an effective way. Recently, Su et al. [356] have proposed a DIV-KCRC method using a bootstrap approach to boost the performance of the base classifiers with the diversity of measures. In another study [357], two methods have been proposed multiview kernel collaborative subspace clustering (MVKCSC) and random

subspace (RS) MVKCSC. Furthermore, a novel approach to forming the Laplacian matrix employing kernel collaborative representation (CR) coefficients is presented for clustering based on subspace clustering and CR theory.

Lv et al. [358] proposed an enhanced ensemble CNN method to classify the imbalanced HSI datasets. First, they used the RS feature method and RF approach. Then, they built an ensemble learning model by integrating the random feature samples and CNN into the HSI classification. In another study, Lv et al. [359] proposed ensemble CNN with feature subspaces (ECNN-EFSs) method, which used an imbalanced training dataset to train the model and acquired higher accuracy. Li et al. [360] proposed a spectral–spatial based feature retrieval approach centered on ensemble empirical mode decomposition for HSI classification. In [361], two novel ensemble-based methods have been proposed, including bagging-based TCRC and boosting-based TCRC approaches, which can increase the accuracy and diversification of the solo classifier method.

Feng et al. [362] proposed an ensemble margin-based semisupervised random forest approach with a limited training dataset. The presented novel approach enables an increase in the efficiency of the ensemble model through adaptively labeling the unlabeled samples with increased classification probability and adding them into the training set. The classification probability of a training sample is reflected by the unsupervised margin value of such samples. The higher the ensemble margin of a sample, the higher the probability of the sample being classified accurately and added into the training set in the next iteration. Liu et al. [363] proposed a novel ensemble self-supervised feature-learning technique that instantly extracts deep features conducive to classification without any annotation details, substantially lowering DL methods' reliance on extensive labeled samples. The first step is to present and use EfficientNet-B0 as the foundation to model input samples in order to more thoroughly and efficiently employ the spatial features in HSIs. Then, the designed model can retrieve the latent features of relatively homogeneous samples grouping together and heterogeneous samples differentiating from each other in a self-supervised form by limiting the cross-correlation matrix of various distortions of the same sample to the identity matrix.

Moreover, they also introduced two novel ensemble approaches, including feature level and view level ensemble. Liu et al. [363] proposed an ensemble stochastic watershed edge weight method. The introduced method consists of two main points: 1) the ensemble approach reduces the variance and 2) categories in HSI datasets and feature similarity have only one-sided implications. Agarwal et al. [364] introduced a novel evolutionary multitask ensemble learning model (EMT EL). First, the model converts the generation of spectral feature subspaces into a multitask optimization problem in order to explore optimized feature subspaces for ensemble learning at the same time, allowing them to choose more useful and representative feature subspaces more efficaciously. Second, obtaining the optimized feature subspace for one core classifier can enable certain other core classification models by accessing valuable features. This can boost descent toward the optimized feature subspace path, minimize luring in locally optimized subspace, and increase exploring potential. Third, randomization-enhanced genetic algorithms are intended for efficient and plausible FE, enabling feature interaction and enhancing the mutual exploring performance of the feature subspace.

Several studies introduce the deep forest (DF) method, which is an ensemble learning approach that has been applied to HSIC. For example, the authors of [365] and [366] used DF for HSIC to acquire enhanced performance. Advanced DF methods have been proposed to tackle the HSIC efficiency problem [367], [368]. These methods integrate the cascade method in DF to boost the performance without facing the computational complexity issue. Cao et al. [369] introduced a novel deep model, named densely connected deep random forest, designed for HSIC. The model comprises multiple forward-connected RFs. Recently, Liu et al. [370] have introduced MAPC-DRF-HSI method for HSIC. The morphological attribute profile cubes undergo scanning with multiple-scale sliding windows to fully exploit the abundant spatial–spectral information. Ultimately, post-multigrained scanning features are fed into a DF classifier to accomplish the classification task.

Shi et al. [371] introduced a promising new extreme-learning-machine-based ensemble transfer learning (TL-ELM) method. TL-ELM acquires the input weights and hidden biases of the ELM learned from the particular domain, as well as samples from the given dataset to recursively alter the output weights of the ELM, which serve as the weights of the building models and, after that, ensemble the training models with their weights for the classification purpose. Liu et al. [372] proposed a novel flexible ensemble method for the multiclass imbalance issue focused on SMOTE and rotation forest (RoF) with diversified sampling rates. The presented method uses SMOTE and a dynamic information sampling ratio for core classifiers to generate a series of balanced datasets with further diversification and much less noise. Feng et al. [373] proposed a label propagation ensemble (LPE) method.

In LPE, an RS technique is presented to split the feature space into various subspaces, and then, several label propagation models are built on correlating subspaces. Eventually, the outcomes from various label propagation models are merged at the classification stage, and only the unlabeled pixels whose label propagation outcomes are identical will be selected with pseudolabels. During the iterative process, ELM models are given training on labeled and pseudolabeled instances. Zhang et al. [374] introduced a multiscale CNNs ensemble-based self-learning (MCE-SL) technique for semisupervised HSI classification. In general, the presented MCE-SL method is divided into two phases. In the first phase, spatial features at various scales are retrieved from sparse labeled training instances in order to train multiple CNN models. The second phase includes classifying the unlabeled instances using the trained multiscale CNNs. After error correction, the problem of labeling partially incorrect is alleviated, and unlabeled samples with high confidence will be added to the original training dataset for the next training iteration.

Fang et al. [375] proposed multiview-based random rotation ensemble pruning and presented some state-of-the-art features. First, employing correlation analysis, the spectral bands are

partitioned into multiple views to ensure that the subsets of spectral bands can adequately obtain the target concept. Second, random rotation, a novel method of space transformation, is presented to transform each view into a variety of coordinate spaces, significantly increasing the diversification of the component classifiers trained on the transformed spaces. Third, an accuracy-guided ensemble pruning approach is formed to eliminate component classifiers with minimal complementarity. As a result, the component classifiers left are put together to produce an ensemble classifier with high complementarity. Zhang et al. [376] presented the RS ensemble with improved feature, which uses improved spatial features to train multiple classifiers. The presented method addresses two basic problems: the curses of imbalanced training samples and a high feature-to-instance ratio. To solve the problem of imbalanced training instances, they first introduce a similar-neighboring-sample search technique. Following that, they generate improved RSs with reduced dimensionality and more unique features than the actual RSs in order to mitigate the curse of a high feature-to-instance ratio more efficiently.

Feng et al. [377] introduced an adaptive semisupervised RoF for HSIC with sparse training data. The proposal relies on RoF, a classification method that has proven to be useful in the case of high-dimensional data. It is altered for SSL by boosting the number of training examples in the learning phase and mining high-quality unlabeled examples with an ensemble margin. The SMOTE method is used to address the issue of class imbalance. In the second step, out-of-bag examples are employed to determine the efficient number of instances to include in the training set. In [378], multiple transferring CNNs are integrated to form a diverse ensemble classification framework.

Furthermore, an enhanced label smoothing method is introduced to enhance the HSI's classification performance. In [379], the vicinity of class noise, a new two-phase ensemble-based data filtering technique is presented to enhance HSIC performance. The presented technique combines noise redundancy classification models with sensitive techniques.

Recently, Lu et al. [380] have introduced two novel dynamic ensemble learning techniques employing local weighted residual (LWR-DEL) and double-weighted residual of multi-CR classifiers. Initially, the dynamic ensemble learning technique focused on clustering is employed to present initial information for the CR classification model. The local weights of every classification model for a distinct area of competence are then acquired using previous information. In order to produce validation instances with global information, the KNN technique is used to take into account the global information of hyperspectral data. The global weights for every classification model can be acquired and applied to limit the locally weighted residuals. Similarly to LWR-DEL, the global information is also utilized to limit the residual, and the last classification performance is then obtained by fusing the double-weighted constrained residuals. Yao et al. [381] introduce a novel multimodal DL framework for LULC classification. They achieve this by making minimal modifications to the conventional vision transformer. Yao et al. [382] investigate the potential of unsupervised cross-modal feature representation learning. They achieve this by incorporating multimodal data into a fully recombined matrix format. Hong et al. [383] present two innovative network components: a self-GAN module and a mutual-GAN module. These additions aim to learn perturbation-insensitive feature representations and bridge the gap between modalities, enhancing effective and robust feature transfer. Zhang et al. [384] propose enhancing FE from LiDAR data using a CNN with an attention mechanism. It introduces a well-designed cascaded block with a short path architecture for efficient multistage information exchange.

I. Future Directions of Ensemble and Deep Ensemble

In the prior section, we discussed the most recent advancements in ensemble and deep ensemble approaches for HSIC. Although ensemble- and deep-ensemble-based HSIC architectures have attracted the attention of the remote sensing community and demonstrated outstanding classification performance, some features, such as the regeneration of sequential input data for ensemble and deep ensemble, demand further investigation. A significant number of the approaches examined are based on the HSI pixel as a sequential point, which is the pixel from each spectral band that makes up a data sequence. While it has the capacity to substantially increase the input sequence for RNNs, this enhancement may give rise to an overfitting issue.

As deep ensemble models continue to demonstrate their efficacy in capturing complex spectral patterns, one key avenue for exploration involves the integration of attention mechanisms to enhance their interpretability and focus on salient features. Researchers may delve into the development of specialized architectures that leverage the inherent characteristics of hyperspectral data, such as its high dimensionality and spectral variability, to further optimize ensemble performance. In addition, investigating the fusion of deep ensembles with multimodal data, such as LiDAR or thermal imagery, could provide a more comprehensive understanding of complex scenes.

Adapting deep ensemble models to operate in real-time and edge computing environments, where efficiency is paramount, represents a critical research direction. As the hyperspectral domain continues to evolve, addressing challenges related to limited labeled data and domain adaptation will be crucial, prompting the exploration of techniques like SSL and TL within the ensemble framework. Moreover, advancing uncertainty quantification methodologies within deep ensembles will contribute to improved model reliability and decision-making transparency. Another interesting future direction may be incorporating EDE-based HSIC architectures in a real-world multitemporal HSI situation.

XII. Experimental Evaluation

Most research-oriented studies presented in the publications introduce a thorough experimental analysis to illustrate the benefits and drawbacks of the proposed methods. However, such studies may have selected various experimental configurations to a certain extent; for example, training, validation, and test samples may have a relatively similar percentage; the samples may be diverse because these samples are ordinarily selected arbitrarily. As a result, in order to compare various studies introduced in the

TABLE I
DETAILED DESCRIPTION OF EACH DATASET USED DURING THE EXPERIMENT

Name	Spatial Dimension	Spectral Bands	Wavelength Range	Classes
IP	145×145	224	400 nm – 2500 nm	16
UP	610×340	103	430 nm – 860 nm	9
SA	512×217	224	360 nm – 2500 nm	16
KSC	610×340	224	400 nm – 2500 nm	13

TABLE II
AMOUNT OF TRAINING AND TEST DATA OF THE SALINAS DATASET

Class No.	Class Name	Training	Test
1	Brocoli_green_weeds_1	201	1808
2	Brocoli_green_weeds_2	373	3353
3	Fallow	198	1778
4	Fallow_rough _plow	140	1254
5	Fallow_smooth	268	2410
6	Stubble	396	3563
7	Celery	358	3221
8	Grapes_untrained	1128	10143
9	Soil_vinyard_develop	621	5582
10	Corn_senesced_green_weeds	328	2950
11	Lettuce_romaine_4wk	107	961
12	Lettuce_romaine_5wk	193	1734
13	Lettuce_romaine_6wk	92	824
14	Lettuce_romaine_7wk	107	963
15	Vinyard_untrained	727	6,541
16	Vinyard_vertical_trellis	181	1626
	Total	5418	48711

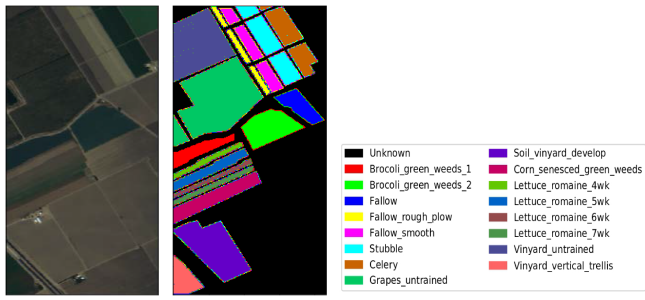
literature comparatively, the experimental configurations must be similar.

Such experimental configurations involve relatively similar samples, and the percentage of samples per round of training in the cross-validation phase should have been chosen. Generally, such samples are selected randomly, and consequently, they are prone to vary for various models if the models are run at various times.

Another problem with the majority of the latest literary works is intersecting between training/test examples, i.e., training/validation examples were chosen arbitrarily for the training and validation step. However, the whole dataset has been carried through a testing step, resulting in a heavily biased model (as the model has already observed the training samples) and increased accuracy. Hence, in this study, the training/test examples are selected arbitrarily (since all of the models were run simultaneously period); moreover, the preceding point was obtained critically, and the correlation between such samples appears void.

A. Experimental Datasets

The first dataset SA was captured by the AVIRIS sensor through the geographical area of Salinas Valley, CA, USA, with a spatial resolution of 3.7 m/pixel [385]. The Salinas scene is made up of 224 spectral bands composed of the wavelength range of 360–2500 nm and 512 × 217 pixels, as shown in Table I. The dataset includes 16 classes made concerning vegetables, vineyard fields, and bare soil. Table II depicts the number of classes and includes training and test data of this dataset. The



(a) (b)

Fig. 14. SA dataset. (a) Source image. (b) Ground truth.

TABLE III
AMOUNT OF TRAINING AND TEST DATA OF THE PAVIA UNIVERSITY DATASET

Class No.	Class Name	Training	Test
1	Asphalt	664	5967
2	Medows	1865	16784
3	Gravel	210	1889
4	Tress	307	2757
5	Metal sheets	135	1210
6	Bare soil	503	4526
7	Bitumen	133	1197
8	Bricks	369	3313
9	Shadows	664	5967
	Total	4850	43610

original and ground truth images of SA data set are shown in the Fig. 14(a) and (b).

The second dataset PU was collected by an unmanned aerial optical device, the Reflective Optics Spectrographic Imaging System over Pavia, Italy, 2002 (ROSIS) [386]. The aircraft was managed by the German Aerospace Center as part of the HySens venture, which was funded by the European Union. After discarding 12 noisy channels, the dataset mainly comprises 640 × 340 pixels with a 1.3-m/pixel spatial resolution and 103 bands containing the wavelength range of 430–860 nm, as shown in Table I. Moreover, the whole dataset is comprised of nine classes. The comprehensive details of different classes are given in Table III. The original and ground truth images of PU data set are shown in the Fig. 15(a) and (b).

The third dataset is the Purdue Indiana Indian Pines scene IP [387], which was acquired from the Indian Pines testing site in North-Western Indiana. The data resolution is 145 × 145, and it incorporates 200 spectral bands. It covers 145 × 145 pixels with a 20-m/pixel spatial resolution and 224 spectral bands comprised from 400 to 2500 nm. The ground truth appends 16 classes of interest, which are largely diverse crops in various growth stages, as given in Tables I and IV. The original and ground truth images of IP data set are shown in the Fig. 16(a) and (b).

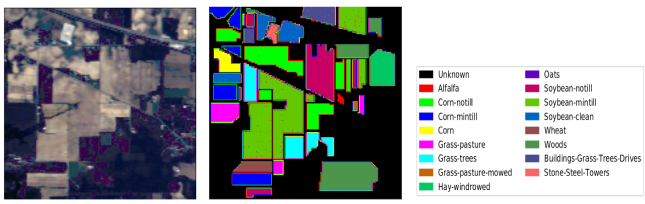
The fourth Kennedy Space Center (KSC) dataset was collected in 1996 by AVIRIS [388], with wavelengths ranging from 400 to 2500 nm. The image of the KSC dataset contains 512 × 614 pixels and 176 spectral bands after the elimination

(a) (b)

Fig. 15. PU dataset. (a) Source image. (b) Ground truth.

TABLE IV
AMOUNT OF TRAINING AND TEST DATA OF THE INDIAN PINES SCENE DATASET

Class No.	Class Name	Training	Test
1	Alfalfa	5	41
2	Corn-notill	143	1285
3	Corn-min	83	747
4	Corn	24	213
5	Grass-pasture	49	434
6	Grass-trees	73	657
7	Grass-pasture-mowed	3	25
8	Hay-windrowed	48	430
9	Oats	2	18
10	Soybean-notill	98	874
11	Soybean-mintill	246	2209
12	Soybean-clean	60	533
13	Wheat	21	184
14	Woods	127	1138
15	Buildings-grass-trees	39	347
16	Stone-steel-towers	10	83
	Total	1031	9218



(a) (b)

Fig. 16. IP dataset. (a) Source image. (b) Ground truth.

of some low-level signal-to-noise ratio bands. The KSC dataset consists of 5202 labeled samples, with a total of 13 upland and wetland classes. The comprehensive details of different classes are given in Table V. Fig. 16 demonstrates the comprehensive class detail and ground truth maps. Furthermore, the disjoint training/test sample maps are shown in Fig. 16(a) and (b). The original and ground truth images of KSC data set are shown in the Fig. 16(a) and (b).

The accompanying tables in this document offer valuable insights into the data used for the subsequent experiments. Table I, to begin with, serves as a comprehensive overview of each dataset employed, providing a succinct yet detailed annotation of their characteristics. On the other hand, Table II documents the count of distinct samples, that is, the train/test samples elected from each class, which have been utilized to achieve all the ensuing experimental outcomes. It is vital to acknowledge the consistency maintained across all experimental procedures, be they our primary focus or the competing methodologies under consideration. Specifically, the quantity of train/test samples, expressed as a percentage, along with the geographical locations from where these train/test samples have been sourced, remains constant across all the experimental techniques. This uniformity ensures a level playing field and a fair comparison of results, thereby enhancing the reliability and validity of our findings.

TABLE V
AMOUNT OF TRAINING AND TEST DATA OF THE KSC SCENE DATASET

Class No.	Class Name	Training	Test
1	Scrub	77	684
2	Willom swamp	25	218
3	CP hammock	26	230
4	Slash pine	26	226
5	Oak/broadleaf	17	144
6	Hardwood	23	206
7	Swap	11	94
8	Graminoid marsh	44	387
9	Spartina marsh	52	468
10	Cattail marsh	41	363
11	Salt marsh	42	377
12	Mud flats	51	452
13	Water	93	834
	Total	528	4683

B. Experimental Results and Discussion

The most significant advancements in recent years have been produced in the areas of HSIC, SVM-RBF[1], CFF-200,[2] 2D-CNN [389], GCNN [225], feedback-attention-based dense convolutional neural network (FADCNN) [390], NL-GCNN [391], DECNN [355], MCE-SL [375], TCRC-Boosting [361], EMT-EL [371], ECNN-PPF-SSF [354], and DHS [392], which have been taken into consideration when comparing the experimental results in order to support the proposals illustrated in this survey. The following are few examples of relevant studies for each of the aforementioned techniques: classification of hyperspectral remote sensing images with SVM-RBF, CCF-200 for HSIC, 2D-CNN, graph-based method for HSIC (GCNN), NL-GCNN, and attention-based CNNs for HSIC (FADCNN), ensemble-based DL based approaches for HSIC (DECNN), multiclass ensemble-based approach for remote sensing image classification (MCE-SL), boosting-based ensemble-based approach for HSIC (TCRC-Boosting), extreme-learning-based ensemble approach for HSIC (EMT-EL), ECNN-PPF-SSF, and DHS.

[1][Online]. Available: https://www.csie.ntu.edu.tw/cjlin/libsvm/
[2][Online]. Available: https://github.com/twgr/ccfs

TABLE VI
CLASSIFICATION ACCURACIES ACQUIRED BY VARIOUS CONVENTIONAL AND STATE-OF-THE-ART METHODS IN TERMS OF OA, AA, AND KAPPA ON THE PAVIA UNIVERSITY DATASET

Class No.	Class names	SVM-RBF	CCF-200	2D CNN	GCNN	FADCNN	NL-GCNN	DECNN	MCE-SL	TCRC-boosting	EMT-EL	E-CNN-PPF-SSF	DHS
1	Asphalt	82.37	86.59	83.85	78.89	99.70	86.80	99.09	98.21	**99.34**	99.25	99.08	99.57
2	Meadows	67.87	72.33	96.09	90.50	99.75	88.74	**100**	**100**	**100**	**100**	**100**	**100**
3	Gravel	69.18	71.75	81.47	71.70	94.31	70.84	98.31	93.25	94.11	94.28	**94.45**	98.43
4	Trees	98.37	99.09	96.12	98.76	99.10	98.43	**100**	99.26	99.26	99.12	99.09	99.50
5	Painted metal sheets	99.41	99.78	98.74	99.93	99.82	99.85	**100**	**100**	**100**	**100**	**100**	**100**
6	Bare Soil	93.64	97.26	49.79	79.08	99.92	94.37	99.16	**100**	99.18	99.42	**100**	**100**
7	Bitumen	91.20	91.88	79.32	71.20	96.97	86.24	**100**	98.42	98.23	99.09	99.09	98.41
8	Self-Blocking Bricks	92.59	94.92	88.89	92.83	97.90	96.74	91.37	94.28	95.41	96.15	96.15	**99.78**
9	Shadows	96.94	98.73	94.19	97.47	98.97	95.78	**100**	99.07	99.03	**100**	99.95	98.53
OA	–	78.89	83.36	86.93	87.08	99.19	90.04	98.61	98.65	98.73	98.80	99.10	**99.49**
AA	–	87.95	90.26	85.38	86.71	98.49	90.87	97.50	97.71	97.99	98.03	98.41	**99.36**
Kappa	–	74.91	79.05	82.42	83.07	95.25	87.06	94.32	93.78	95.21	94.85	95.37	**99.31**

TABLE VII
CLASSIFICATION ACCURACIES ACQUIRED BY VARIOUS CONVENTIONAL AND STATE-OF-THE-ART METHODS IN TERMS OF OA, AA, AND KAPPA ON THE SALINAS DATASET

Class No.	SVM-RBF	CCF-200	2D CNN	GCNN	FADCNN	NL-GCNN	DECNN	MCE-SL	TCRC-boosting	EMT-EL	E-CNN-PPF-SSF	DHS
1	98.98	99.49	71.57	99.59	88.26	99.69	**100**	**100**	**100**	**100**	**100**	**100**
2	99.67	99.95	99.86	98.07	98.04	99.21	**100**	**100**	**100**	**100**	**100**	**100**
3	98.70	99.43	88.89	91.95	98.04	99.79	**100**	**100**	**100**	**100**	**100**	**100**
4	97.77	99.33	98.14	97.84	97.13	98.29	**100**	**100**	**100**	**100**	**100**	**100**
5	98.33	98.82	98.17	98.06	99.13	99.28	**100**	**100**	99.28	99.42	**100**	99.97
6	99.72	99.80	**100**	99.00	99.10	99.80	**100**	**100**	**100**	**100**	**100**	**100**
7	99.46	99.66	97.00	99.29	99.21	99.04	**100**	**100**	**100**	**100**	**100**	99.97
8	70.37	67.56	70.79	82.25	75.72	79.11	96.14	97.35	**98.35**	98.33	98.21	99.98
9	98.59	99.19	99.45	97.11	**100**	97.74	**100**	**100**	**100**	**100**	**100**	**100**
10	93.74	93.80	96.19	91.60	79.00	95.01	**100**	99.11	99.34	99.12	99.33	99.93
11	94.70	95.87	96.37	90.77	95.3	94.60	99.43	**100**	**100**	99.17	**100**	**100**
12	99.89	99.95	**100**	**100**	98.34	**100**	**100**	**100**	**100**	**100**	**100**	**100**
13	97.81	98.15	**100**	98.96	95.12	98.96	**100**	**100**	**100**	**100**	**100**	**100**
14	97.35	96.86	98.33	97.35	98.86	99.41	**100**	**100**	**100**	99.43	99.37	**100**
15	71.53	80.77	91.22	70.44	99.23	84.26	98.01	**98.40**	98.23	98.39	97.23	99.80
16	98.18	98.18	93.00	97.10	83.53	98.01	**100**	**100**	**100**	**100**	**100**	**100**
OA	88.82	89.72	90.25	90.37	90.58	92.28	98.84	99.08	99.10	99.17	99.18	**99.96**
AA	94.67	95.43	93.69	94.34	94.56	96.39	99.37	99.54	99.45	99.45	99.58	**99.97**
K	87.57	88.58	89.18	89.28	90.25	89.64	97.15	98.23	97.56	98.11	98.17	**99.95**

TABLE VIII
CLASSIFICATION ACCURACIES ACQUIRED BY VARIOUS CONVENTIONAL AND STATE-OF-THE-ART METHODS IN TERMS OF OA, AA, AND KAPPA ON THE INDIAN PINES DATASET

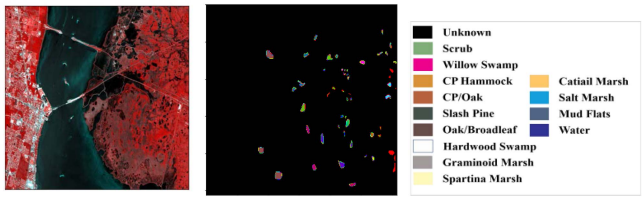
Class No.	SVM-RBF	CCF-200	2D CNN	GCNN	FADCNN	NL-GCNN	DECNN	MCE-SL	TCRC-boosting	EMT-EL	E-CNN-PPF-SSF	DHS
1	71.39	76.37	54.77	76.66	88.26	83.09	89.06	88.17	89.75	90.21	**91.35**	98.01
2	71.05	77.93	96.94	86.10	98.04	89.03	**98.14**	97.03	97.39	97.01	97.51	95.50
3	86.96	94.57	99.46	100	97.04	100	97.74	98.16	98.36	98.17	**98.91**	95.02
4	91.72	94.41	**96.87**	93.06	96.03	93.51	96.15	96.37	96.83	96.56	96.60	97.32
5	85.80	91.39	94.12	92.06	99.34	94.12	99.44	99.01	**99.42**	99.30	98.03	98.10
6	93.85	97.04	96.81	96.81	**99.48**	98.18	99.17	99.31	98.43	98.68	98.72	99.53
7	75.38	90.96	**91.29**	88.24	75.72	88.24	75.89	76.94	77.94	77.55	78.85	**100**
8	59.88	69.48	93.05	76.80	**100**	78.78	**100**	**100**	**100**	**100**	**100**	**100**
9	76.24	**89.01**	87.59	80.85	78.00	86.70	78.19	79.05	79.44	79.51	80.21	92.30
10	96.91	98.77	**100**	99.38	96.50	99.38	96.09	96.37	96.85	96.62	96.78	96.12
11	79.58	93.73	68.57	93.89	98.49	94.94	98.01	98.49	99.03	**100**	**100**	98.64
12	74.84	74.55	88.48	93.64	95.45	**97.27**	95.39	96.09	96.37	96.08	96.71	94.56
13	97.78	**100**	**100**	**100**	99.02	**100**	**100**	**100**	**100**	99.93	**100**	**100**
14	79.49	97.44	82.05	92.31	**99.48**	97.44	99.37	99.09	98.97	99.05	99.25	99.78
15	**100**	90.91	**100**	**100**	99.74	**100**	99.62	99.89	98.91	98.64	98.34	98.70
16	**100**	**100**	**100**	**100**	82.80	**100**	83.09	84.34	85.55	85.94	86.07	90.60
OA	74.24	82.87	90.25	85.43	94.80	87.92	95.71	95.89	95.93	95.96	96.61	**97.91**
AA	83.80	89.78	93.69	91.87	93.96	93.79	94.96	95.01	95.24	95.41	95.57	**97.14**
Kappa	70.93	80.59	89.18	83.42	86.10	86.25	88.50	89.23	90.06	91.32	91.96	**96.26**

All the aforementioned research studies are focused on CNN, graph-based, and ensemble-based frameworks and are analyzed on four benchmark HSI datasets, notably IP, PU, KSC, and SA scene. This survey focuses solely on the effectiveness of all these models while taking into account the relatively small sample of training datasets used to classify HSI for joint spatial–spectral classification.

The experimental results are mentioned here, along with a thorough analysis of the research results. Tables VI–IX and Figs. 18–21 demonstrate the acquired accuracies for disjoint training and test samples. The ten-cross-validation operation was used to analyze the overall accuracy (OA%), average accuracy (AA%), and Kappa (κ) accuracy for comparative processes in all of the results depicted in Tables VI–IX and Figs. 18–21. For this case specifically, let us suppose the case of Salinas scene results; the work [392] has the highest OA, AA, and κ accuracies, which are 99.96%, 99.97%, and 99.95%, respectively, in comparison with the OA, AA, and κ accuracies for ensemble-based methods comparative results; 99.18%, 99.58%, and 98.17% for [354], 99.17%, 99.45%, and 98.11% for [371], 99.10%, 99.45%, and 97.56% for [361], 99.08%, 99.54%, and 98.23% for [375], 98.84%, 99.37%, and 97.15% for [355], and

TABLE IX
CLASSIFICATION ACCURACIES ACQUIRED BY VARIOUS CONVENTIONAL AND STATE-OF-THE-ART METHODS IN TERMS OF OA, AA, AND KAPPA ON THE KSC SCENE DATASET

Class No.	SVM-RBF	CCF-200	2D CNN	GCNN	FADCNN	NL-GCNN	DECNN	MCE-SL	TCRC-Boosting	EMT-EL	ECNN-PPF-SSF	DHS
1	95.10	96.23	**100**	90.14	97.33	72.34	**100**	**100**	96.15	**100**	**100**	**100**
2	62.14	97.21	85.67	91.31	61.25	93.41	99.13	73.34	99.25	97.31	**100**	**100**
3	89.12	71.31	73.59	73.24	66.37	81.44	92.27	85.16	98.38	**100**	**100**	**100**
4	74.19	66.15	73.45	71.35	68.91	73.87	**99.35**	85.29	87.32	64.22	67.34	85.34
5	69.23	00.69	93.19	81.42	65.16	90.65	**100**	**100**	99.47	98.18	**100**	**100**
6	53.21	00.24	98.51	25.23	83.89	98.22	**100**	87.43	95.43	**100**	98.16	97.33
7	61.25	00.00	99.51	88.92	81.67	99.42	56.41	95.16	**100**	76.07	**100**	**100**
8	97.29	78.42	84.33	79.33	64.88	71.29	65.71	99.11	**100**	97.15	**100**	**100**
9	98.35	99.34	97.28	92.35	**100**	**100**	**100**	**100**	**100**	**100**	99.33	99.15
10	50.33	**100** 99.12	59.15	89.37	92.66	**100**	**100**	**100**	82.18	99.43	99.41	98.27
11	00.00	**100**	98.24	70.37	97.68	90.43	99.16	**100**	**100**	**100**	**100**	**100**
12	76.38	89.89	12.43	83.98	94.27	99.12	**100**	**100**	**100**	**100**	**100**	**100**
13	98.43	82.91	99.28	**100**	**100**	**100**	**100**	**100**	97.42	**100**	**100**	**100**
OA	78.53	83.81	84.72	85.56	87.84	89.57	93.52	96.09	96.16	96.21	97.36	**98.76**
AA	72.26	69.86	84.31	78.92	85.26	90.69	93.41	94.05	95.64	94.67	95.33	**98.46**
K	76.04	81.83	83.02	83.82	86.51	88.44	92.80	95.64	95.72	95.78	97.06	**98.58**



(a) (b)

Fig. 17. KSC dataset. (a) Source image. (b) Ground truth.

for nonensemble methods are 92.28%, 96.39%, and 89.64% for [391], 90.58%, 94.56%, and 89.28% for [390], 90.25%, 93.69%, and 89.18% for [225], 89.72%, 95.43%, and 88.58% for CFF-200, and 88.82%, 94.67%, and 87.57% for SVM-RBF. Similar findings were produced with the other experimental datasets. Tables VI–IX highlight the increased accuracies in bold.

As illustrated in Table II and Fig. 18, the comparison techniques mainly misclassify samples with relatively comparable structures (i.e., Grapes_untrained, Corn_senesced_green_weeds, and Vinyard_untrained classes for the SA scene dataset). Furthermore, the overall accuracy for Grapes Untrained is relatively low compared to the other classes because of the factors stated above. In simple terms, greater accuracy can be obtained by enhancing the number of labeled training samples. As a result, a greater proportion of labeled training samples can generate increased accuracy scores for all comparative techniques.

Overall, the techniques in [392], [354], and [371] performed better (i.e., produced reliable performance) than the other similar techniques, particularly when there were fewer training samples with labels. The aforementioned information gives rise to the conclusion that the quantity of training samples has no effect on such works. In addition, the accuracy of such techniques improves with the number of training samples. However, in comparison to these techniques, the rest of the techniques can perform more effectively with a larger number of training samples. The same pattern has been demonstrated with a larger number of training samples. Thereby, analyzing disjoint train/test samples, the studies [392], [354], and [371] can resolve the problem of the lack of presence of training samples.

Furthermore, we can deduce that the CNN-based models do not outperform the ensemble models alone, despite the fact that unsupervised techniques do not need samples to be labeled; when there are no limitations, such techniques may learn nothing. In addition, the CNN has a symmetric architecture that also results in excess training parameters, increasing the complexity of training. The studies [391] and [390] mitigate the aforementioned problems. The research study [225] does not use the greedy layerwise strategy, resulting in poor performances; hence, such techniques have a gap for improved performance.

In fact, due to the scarcity of labeled training instances, CNN-based classification outperforms classical machine learning techniques. Whereas CNNs can learn the inner architecture of unlabeled data, the final feature interpretation may lack task-driven properties. Furthermore, the graph-based CNN impacts the choice of the most essential samples for training, allowing the model to emphasize more on identical samples for HSIC, although FSL enhances the interpretation of sample relationships to obtain a discriminative decision boundary for HSIC. TL supports utilization correlation among various HSIs to help mitigate the amount of training needed as well as the number of trainable parameters while improving model effectiveness. When analyzing HSI using only the raw data (i.e., without retrieving or learning the features), DA produces enough samples, increasing the variety of samples.

C. *Experimental Analysis of the CNN*

This section covers some deep CNN hyperspectral FE methods: typical convolutional method to a state-of-the-art NL-GCNN method. Moreover, we performed a series of experimental studies utilizing various cutting-edge research studies published in recent years. This experimental study is intended to evaluate the efficiency of the CNN procedures instead of the model's performance. The standard models employ fully CNN FE methods, such as 2-D CNN for HSIC [389], GCNN for HSIC [225], and, lately, FADCNN [390], to retrieve the fine-grained spectral–spatial feature representation. In order to retrieve the intensity-level semantic features and the gradient-level features from the HSIs, the GCNN process uses a fusion of graph-based methods.

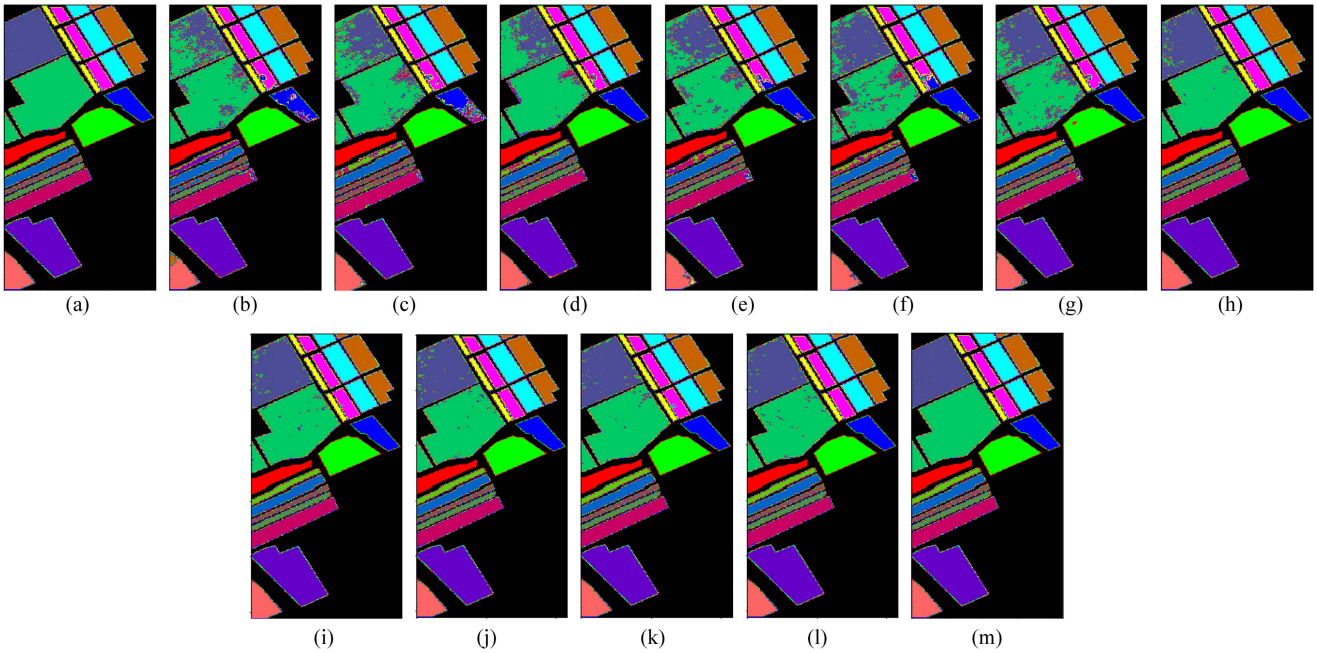


Fig. 18. Classification maps of various methods for the Salinas dataset. From left to right: (a) ground truth, (b) SVM-RBF, (c) CCF-200, (d) 2-D CNN, (e) GCNN, (f) FADCNN, (g) NL-GCNN, (h) DECNN, (i) MCE-SL, (j) TCRC-Boosting, (k) EMT-EL, (l) ECNN-PPF-SSF, and (m) DHS.

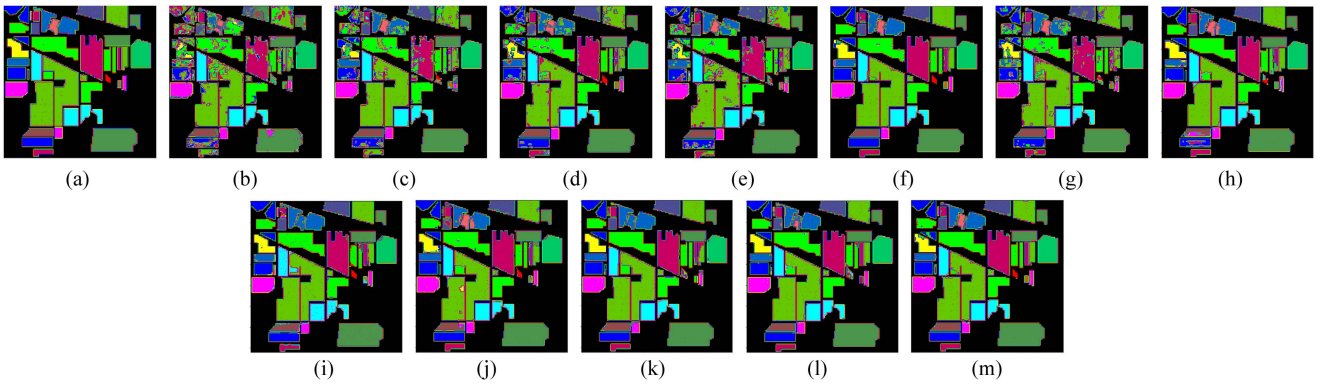


Fig. 19. Classification maps of various methods for the Indian Pines dataset. From left to right and from top to bottom: (a) ground truth, (b) SVM-RBF, (c) CCF-200, (d) 2-D CNN, (e) GCNN, (f) FADCNN, (g) NL-GCNN, (h) DECNN, (i) MCE-SL, (j) TCRC-Boosting, (k) EMT-EL, (l) ECNN-PPF-SSF, and (m) DHS.

All the previously mentioned CNN FE methods have been tested on four various hyperspectral datasets, such as the IP, PU, SA, and KSC datasets. Tables VI–IX demonstrate the experimental accuracy. From all these analyses, it is evident that the FADCNN method showed better performance than 2-D CNN, GCNN, and NL-GCNN. All the datasets have shown a typical pattern, where the GCNN and NL-GCNN methods outperformed the conventional 2-D CNN and classical machine learning methods SVM-RBF and CCF-200 processes, respectively. Comparing the performance variation to the NL-GCNN for other datasets, it showed significant improvement against the rest of CNN and classical machine learning methods. Notably, the stability and classification accuracy of conventional CNNs can be improved by adding the 2-D CNN to the 3-D CNN convolution process, which is easy to execute and plug into.

XIII. Conclusion and Future Work

A. Conclusion

The rich details stored in HSI data are an enticing feature that contributes to the use of HSI technology in real-world applications. Furthermore, advancements in machine learning techniques boost the integration opportunity of such tools. In this article, we surveyed novel approaches in HSIC utilizing cutting-edge DNNs (for example, AE, DBN, RNN, CNN, TL, FSL, AL/self-learning, and DA), ensemble, and deep ensemble in a wide range of learning approaches. Moreover, we explored techniques for overcoming the difficulties of the sparsity of training data accessibility, such as DA, FSL, TL, and AL. We chose several comparative studies to perform experimentations on benchmark HSI datasets using the techniques mentioned in the previous section.

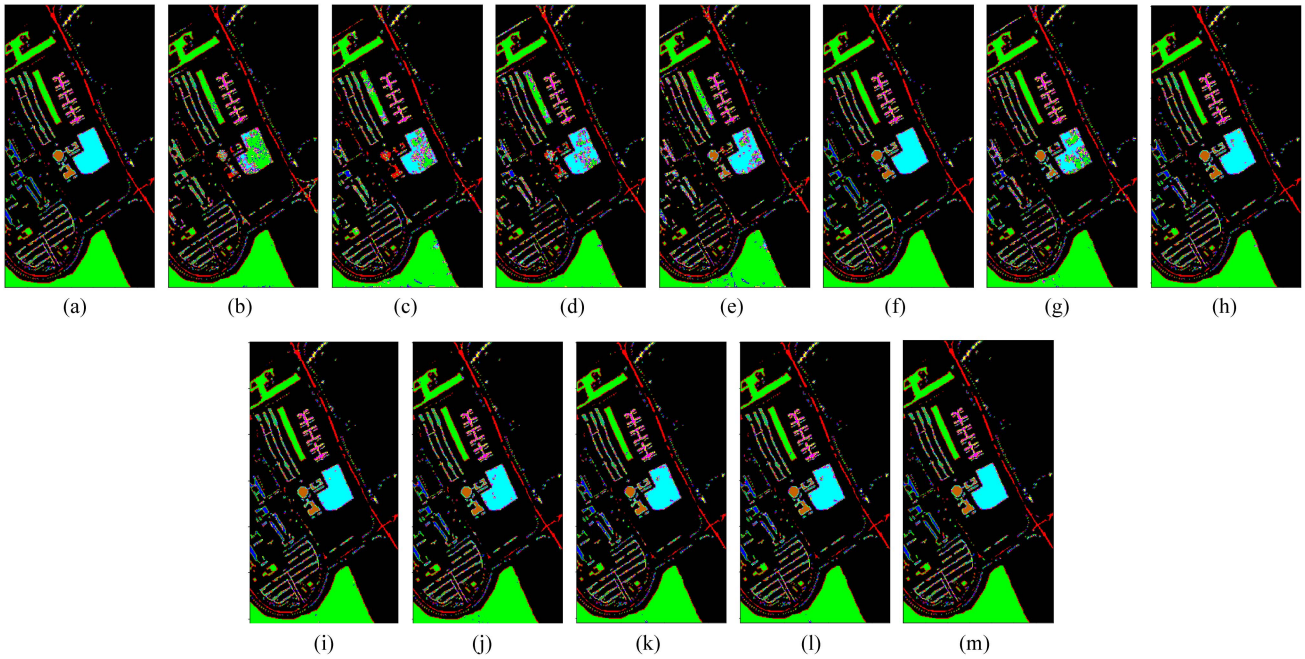


Fig. 20. Classification maps of various methods for the Pavia University dataset. From left to right: (a) ground truth, (b) SVM-RBF, (c) CCF-200, (d) 2-D CNN, (e) GCNN, (f) FADCNN, (g) NL-GCNN, (h) DECNN, (i) MCE-SL, (j) TCRC-Boosting, (k) EMT-EL, (l) ECNN-PPF-SSF, and (m) DHS.

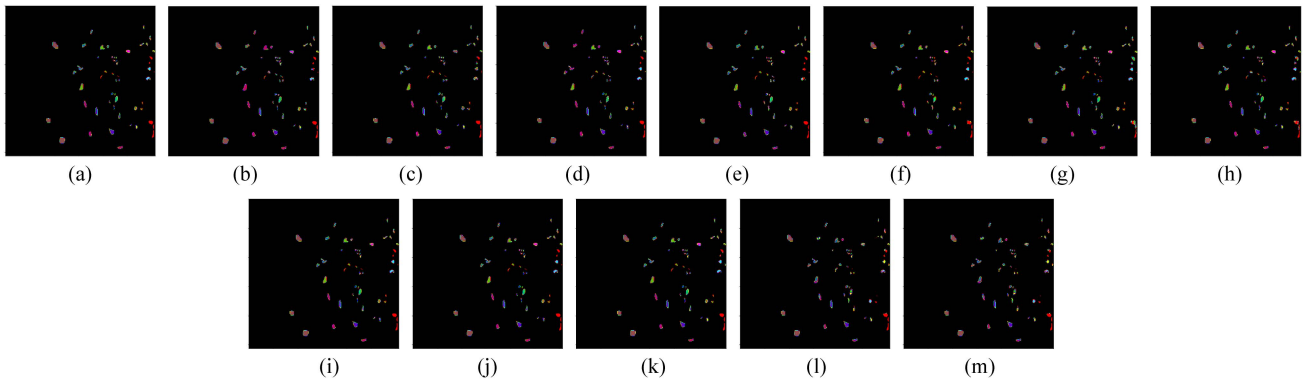


Fig. 21. Classification maps of various methods for the KSC dataset. From left to right: (a) ground truth, (b) SVM-RBF, (c) CCF-200, (d) 2-D CNN, (e) GCNN, (f) FADCNN, (g) NL-GCNN, (h) DECNN, (i) MCE-SL, (j) TCRC-Boosting, (k) EMT-EL, (l) ECNN-PPF-SSF, and (m) DHS.

B. Future Work

While such HSIC methods represent a task that is fast, remarkable, and sophisticated, additional research is needed to boost generalization performance.

1) The primary problem with DNN-based HSIC is a scarcity of labeled data. Because of the scarcity of labeled data, HSI data are notorious, and DNNs require a significant percentage of labeled training data. Section XI explored several extensively utilized approaches for dealing with the aforementioned problem, but substantial progress is still required to make effective utilization of the scarcity of training data. One solution to this issue could be to evaluate the incorporation of different learning approaches mentioned in Section XI in order to capitalize on the joint advantage. Another strategy is to use a few-shot or K-shot learning method, which can effectively predict class labels with only a few labeled examples.
2) Furthermore, it is essential to contribute to the joint exploitation of spectral–spatial HSI features to support classification performance acquired by the previously stated HSIC methodologies. Computationally effective and robust frameworks are also a promising choice for HSIC. Hence, the problem of the increased computational complexity of DNNs is of utmost significance, and it is vital to incorporate parallel HSIC frameworks to boost the execution of DNNs to reach the computational requirements of computationally expensive HSI applications. In such a prospect, the parallel HSIC approaches can be implemented using HPC platforms and highly specialized physical devices like GPUs and FPGAs. Therefore, incorporating the aforementioned factors into building a novel HSIC

approach requires adequately utilizing the insufficient training samples while analyzing joint spectral–spatial features of HSI and sustaining a minimum computational complexity.

3) Another future perspective, adversarial examples or machine illusions, has garnered substantial attention within the DL community. They can be considered part of the DA family, involving the injection of systematic noise into an image. This injection of noise can lead to a markedly different prediction by the CNN model. Adversarial training involves incorporating the examples into the training set, enhancing the model's resilience against attacks. Given their ability to expose vulnerabilities in a trained model, the use of adversarial examples as a DA strategy proves to be an effective approach.
4) Knowledge distillation has been widely explored in the context of natural image classification, and its application to HSIC has been an emerging area of research. Knowledge distillation involves training a smaller model (student) to mimic the behavior and predictions of a larger more complex model (teacher).
5) Incremental learning is a dynamic paradigm in machine learning that enables models to adapt and evolve over time as new data become available. Incremental learning holds significant potential for HSIC, providing a framework for continuous adaptation and improvement as new hyperspectral data become available. In the context of HSI, where each pixel in an image is associated with a spectrum of information across multiple bands, incremental learning can offer several advantages.

REFERENCES

[1] J. T. Rayner, "SpeX: A medium-resolution 0.8–5.5 micron spectrograph and imager for the NASA infrared telescope facility," *Pub. Astron. Soc. Pacific*, vol. 115, no. 805, pp. 362–382, 2003.

[2] M. Ahmad, "Spatial prior fuzziness pool-based interactive classification of hyperspectral images," *Remote Sens.*, vol. 11, no. 9, 2019, Art. no. 1136.

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

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

[5] M. H. Khan, "Hyperspectral imaging for color adulteration detection in red chili," *Appl. Sci.*, vol. 10, no. 17, 2020, Art. no. 5955.

[6] 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.

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

[8] 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, 2020, Art. no. 7783.

[9] M. H. Khan, "Hyperspectral imaging-based unsupervised adulterated red chili content transformation for classification: Identification of red chili adulterants," *Neural Comput. Appl.*, vol. 33, no. 21, pp. 14507–14521, 2021.

[10] "10 important applications of hyperspectral image." Accessed: May 30, 2023. [Online]. Available: https://grindgis.com/remote-sensing/10-important-applications-of-hyperspectral-image

[11] 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.

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

[13] "hyperspectral image classification application." Accessed: May 30, 2023. [Online]. Available: https://resonon.com/applications

[14] 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.

[15] 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, pp. 439–462, 2018.

[16] D. Hong, N. Yokoya, J. Chanussot, J. Xu, and X. 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.

[17] M. Ahmad, S. Lee, I. U. Haq, and Q. Mushtaq, "Hyperspectral remote sensing: Dimensional reduction and end member extraction," *Int. J. Soft Comput. Eng.*, vol. 2, no. 6, pp. 170–175, 2012.

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

[19] 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.

[20] J. M. Bioucas-Dias, "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.

[21] M. Ahmad, A. K. Bashir, and A. M. Khan, "Metric similarity regularizer to enhance pixel similarity performance for hyperspectral unmixing," *Optik*, vol. 140, pp. 86–95, 2017.

[22] 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.

[23] M. Ahmad and I. U. Haq, "Linear unmixing and target detection of hyperspectral imagery using OSP," *Proc. IPCSIT*, vol. 10, pp. 179–183, 2011.

[24] M. Ahmad, D. I. U. Haq, 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, no. 6, pp. 606–614, 2011.

[25] M. Ahmad, I. Ulhaq, and Q. Mushtaq, "Aik method for band clustering using statistics of correlation and dispersion matrix," in *Proc. Int. Conf. Inf. Commun. Manage.*, 2011, vol. 10, pp. 114–118.

[26] 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.

[27] 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.

[28] 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.

[29] 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.

[30] 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.

[31] 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.

[32] 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.

[33] 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.

[34] P. Chen, B. Zhang, D. Hong, Z. Chen, X. Yang, and B. Li, "FCCDN: Feature constraint network for VHR image change detection," *ISPRS J. Photogrammetry Remote Sens.*, vol. 187, pp. 101–119, 2022.

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

[36] 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, no. 1, 2018, Art. no. e0188996.

[37] 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*, vol. 206, 2020, Art. no. 163712.

[38] 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.

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

[40] 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*, vol. 180, pp. 370–378, 2019.

[41] 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, 2021.

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

[43] 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.

[44] 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.

[45] 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.

[46] 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.

[47] 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.

[48] 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.

[49] 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, 2016, Art. no. 436.

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

[51] D. Hong, N. Yokoya, N. Ge, J. Chanussot, and X. 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.

[52] 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.

[53] 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.

[54] 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.

[55] 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.

[56] 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.

[57] S. B. Kotsiantis, I. Zaharakis, and P. Pintelas, "Supervised machine learning: A review of classification techniques," *Emerg. Artif. Intell. Appl. Comput. Eng.*, vol. 160, no. 1, pp. 3–24, 2007.

[58] F. Ullah, B. Zhang, and R. U. Khan, "Image-based service recommendation system: A JPEG-coefficient RFs approach," *IEEE Access*, vol. 8, pp. 3308–3318, 2019.

[59] 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.

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

[61] 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.

[62] 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.

[63] 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.

[64] 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.

[65] P. Ghamisi, "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.

[66] 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, 2019, Art. no. 204.

[67] B. Rasti, "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.

[68] 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.

[69] 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.*, vol. 19, 2020, Art. no. 5502205.

[70] D. Hong, "SpectralFormer: Rethinking hyperspectral image classification with transformers," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5518615.

[71] 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.

[72] 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.

[73] 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, 2016, Art. no. 483.

[74] 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.

[75] A. Oliva and A. Torralba, "Modeling the shape of the scene: A holistic representation of the spatial envelope," *Int. J. Comput. Vis.*, vol. 42, pp. 145–175, 2001.

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

[77] 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.

[78] 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.

[79] B. Zhang, 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, no. 1, pp. 1–15, 2012.

[80] 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.

[81] 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.

[82] R. Azhar, D. Tuwohingide, D. Kamudi, and N. Suciati, "Batik image classification using SIFT feature extraction, bag of features and support vector machine," *Procedia Comput. Sci.*, vol. 72, pp. 24–30, 2015.

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

[84] 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, vol. 2, pp. 704–708.

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

[86] 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.

[87] 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.

[88] 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.

[89] 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.

[90] 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.

[91] 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.

[92] 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.

[93] 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. 2009.

[94] 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.

[95] 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.

[96] 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.

[97] 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, 2016, Art. no. 157.

[98] 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.

[99] 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.

[100] 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.

[101] 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. 083690.

[102] 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.

[103] 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.

[104] 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, 2016, Art. no. 259.

[105] 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.

[106] 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.

[107] 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.

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

[109] D. Hong, "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.

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

[111] 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.

[112] 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.

[113] C. Li, B. Zhang, D. Hong, J. Yao, and J. Chanussot, "LRR-Net: An interpretable deep unfolding network for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5513412.

[114] S. Chen and Y. Wang, "Convolutional neural network and convex optimization," Dept. Elect. Comput. Eng., Univ. California at San Diego, San Diego, CA, USA, Tech. Rep, 2014.

[115] P. Hammer, "Adaptive control processes: A guided tour (R. Bellman)," *Math. Gazette*, vol. 46, no. 356, pp. 160–161, 1962.

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

[117] M. Reichstein, G. Camps-Valls, B. Stevens, M. Jung, J. Denzler, and N. Carvalhais, "Deep learning and process understanding for data-driven earth system science," *Nature*, vol. 566, no. 7743, pp. 195–204, 2019.

[118] 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.

[119] Q. Nguyen and M. Hein, "Optimization landscape and expressivity of deep CNNs," in *Proc. Int. Conf. Mach. Learn.*, 2018, pp. 3730–3739.

[120] 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.

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

[122] L. Bottou, "Stochastic gradient learning in neural networks," *Proc. Neuro-Nimes*, vol. 91, no. 8, p. 12, 1991.

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

[124] G. Hinton, N. Srivastava, and K. Swersky, "Neural networks for machine learning: Lecture 6a—Overview of mini-batch gradient descent," *Cited on*, vol. 14, no. 8, p. 2, 2012.

[125] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.

[126] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," 2017, *arXiv:1711.05101*.

[127] 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.

[128] L. Liu et al., "On the variance of the adaptive learning rate and beyond," in *Proc. 8th Int. Conf. Learn. Representations*, Apr. 2020.

[129] H. Yong, J. Huang, X. Hua, and L. Zhang, "Gradient centralization: A new optimization technique for deep neural networks," in *Proc. 16th Eur. Conf. Comput. Vis.*, 2020, pp. 635–652.

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

[131] D. Erhan, A. Courville, Y. Bengio, and P. Vincent, "Why does unsupervised pre-training help deep learning?," in *Proc. 13th Int. Conf. Artif. Intell. Statist. Workshop Conf. Proc.*, 2010, pp. 201–208.

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

[133] 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.

[134] 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.

[135] 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.

[136] 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.

[137] Y. Fang, "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.

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

[139] 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.

[140] 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.

[141] 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.

[142] 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.

[143] 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.

[144] 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.

[145] 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.

[146] 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.

[147] 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.

[148] 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, 2018, Art. no. 1454.

[149] 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, 2018, Art. no. 395.

[150] 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, 2018, Art. no. 2831.

[151] 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.

[152] 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.

[153] 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.

[154] J. Acquarelli, E. Marchiori, L. M. 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, 2018, Art. no. 1156.

[155] 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, 2017, Art. no. 1330.

[156] 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.

[157] F. Ullah, "Deep edu: A deep neural collaborative filtering for educational services recommendation," *IEEE Access*, vol. 8, pp. 110915–110928, 2020.

[158] F. Ullah, B. Zhang, R. U. Khan, I. Ullah, A. Khan, and A. M. Qamar, "Visual-based items recommendation using deep neural network," in *Proc. Int. Conf. Comput., Netw. Internet Things*, 2020, pp. 122–126.

[159] 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.

[160] 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.

[161] 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, 2020, Art. no. 159.

[162] Z. Zhang, "Semi-supervised hyperspectral image classification algorithm based on graph embedding and discriminative spatial information," *Microprocessors Microsyst.*, vol. 75, 2020, Art. no. 103070.

[163] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, 2017.

[164] L. Liu, C. Shen, and A. Van Den Hengel, "The treasure beneath convolutional layers: Cross-convolutional-layer pooling for image classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 4749–4757.

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

[166] S. K. Roy, S. Manna, S. R. Dubey, and B. B. Chaudhuri, "LiSHT: Non-parametric linearly scaled hyperbolic tangent activation function for neural networks," in *Proc. Int. Conf. Comput. Vis. Image Process.*, 2022, pp. 462–476.

[167] T. Williams and R. Li, "Wavelet pooling for convolutional neural networks," in *Proc. Int. Conf. Learn. Represent.*, 2018.

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

[169] 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.

[170] 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.

[171] 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.

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

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

[174] 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, 2018, Art. no. 349.

[175] W. Zhao, "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.

[176] 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.

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

[178] 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.

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

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

[181] 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, 2018, Art. no. 212.

[182] 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.*, 2018, pp. 1–4.

[183] 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.*, 2018, pp. 171–179.

[184] G. Turra, S. Arrigoni, and A. Signoroni, "CNN-based identification of hyperspectral bacterial signatures for digital microbiology," in *Proc. 19th Int. Conf. Image Anal. Process.*, 2017, pp. 500–510.

[185] 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.

[186] 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.

[187] 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.

[188] 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, 2022.

[189] 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, 2017, Art. no. 618.

[190] 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.

[191] 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.

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

[193] 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.

[194] M. E. Paoletti, J. M. 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.

[195] 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.

[196] S. K. Roy, M. E. Paoletti, J. M. Haut, E. M. 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.

[197] 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.

[198] 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.

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

[200] 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.

[201] 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.

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

[203] 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.

[204] 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.

[205] 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.

[206] 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.

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

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

[209] 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.

[210] 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.

[211] 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.

[212] L. Liu, "Multiscale deep spatial feature extraction using virtual RGB image for hyperspectral imagery classification," *Remote Sens.*, vol. 12, no. 2, 2020, Art. no. 280.

[213] 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.

[214] 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. With Appl.*, vol. 129, pp. 246–259, 2019.

[215] 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.

[216] 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.

[217] 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.*, vol. 19, 2022, Art. no. 5509705.

[218] W. Song, S. Li, L. Fang, and T. Lu, "Hyperspectral image classification with deep feature fusion network," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 6, pp. 3173–3184, Jun. 2018.

[219] 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.

[220] 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.

[221] 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, 2022, Art. no. 2501605.

[222] 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.*, vol. 60, 2022, Art. no. 5516619.

[223] W. Song, Y. Dai, Z. Gao, L. Fang, and Y. Zhang, "Hashing-based deep metric learning for the classification of hyperspectral and LiDAR data," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5704513.

[224] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," in *Proc. Int. Conf. Learn. Representations*, 2017.

[225] 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.

[226] 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.

[227] 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.

[228] 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.

[229] A. Hassanzadeh, A. Kaarna, and T. Kauranne, "Unsupervised multi-manifold classification of hyperspectral remote sensing images with contractive autoencoder," in *Proc. 20th Scand. Conf. Image Anal.*, 2017, 2017, pp. 169–180.

[230] Y. Wang, Y. Jiang, Y. Wu, and Z.-H. Zhou, "Multi-manifold clustering," in *Proc. 11th Pacific Rim Int. Conf. Artif. Intell.*, 2010, pp. 280–291.

[231] 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.*, 2011, pp. 833–840.

[232] 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.

[233] 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.

[234] 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.

[235] 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.

[236] 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.

[237] 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.

[238] 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.

[239] M. Ahmad, A. M. Khan, M. Mazzara, and S. Distefano, "Multi-layer extreme learning machine-based autoencoder for hyperspectral image classification," in *Proc. Int. Joint Conf. Comput. Vis., Imag. Comput. Graph. Theory Appl.*, 2019, pp. 75–82.

[240] 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.

[241] 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.

[242] 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.

[243] B. Liu, Q. Zhang, Y. Li, 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, pp. 1989–1997, 2019.

[244] 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.

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

[246] B. Ayhan and C. Kwan, "Application of deep belief network to land cover classification using hyperspectral images," in *Proc. 14th Int. Symp. Adv. Neural Netw.*, 2017, pp. 269–276.

[247] 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*.

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

[249] H. Xiong, A. J. Rodríguez-Sánchez, S. Szedmak, and J. Piater, "Diversity priors for learning early visual features," *Front. Comput. Neurosci.*, vol. 9, 2015, Art. no. 104.

[250] 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.

[251] 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, 2018, Art. no. 396.

[252] 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.

[253] 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.

[254] 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.

[255] 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.

[256] 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.

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

[258] 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.

[259] 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.

[260] 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. 1st Chin. Conf. Pattern Recognit. Comput. Vis.*, 2018, pp. 409–420.

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

[262] 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.

[263] M. Seydgar, A. Alizadeh 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, 2019, Art. no. 883.

[264] 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.

[265] 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.

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

[267] 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.

[268] 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.

[269] 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.

[270] 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.

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

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

[273] 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.

[274] 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.

[275] 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, 2018, Art. no. 574.

[276] 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.

[277] 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.

[278] 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.

[279] P. Bateni, R. Goyal, V. Masrani, F. Wood, and L. Sigal, "Improved few-shot visual classification," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 14493–14502.

[280] B. Liu, X. Yu, A. Yu, P. Zhang, G. Wan, and R. Wang, "Deep few-shot learning for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 4, pp. 2290–2304, Apr. 2019.

[281] Y. Zhang, W. Li, M. Zhang, S. Wang, R. Tao, and Q. Du, "Graph information aggregation cross-domain few-shot learning for hyperspectral image classification," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, doi: 10.1109/TNNLS.2022.3185795.

[282] N. Li, D. Zhou, J. Shi, X. Zheng, T. Wu, and Z. Yang, "Graph-based deep multitask few-shot learning for hyperspectral image classification," *Remote Sens.*, vol. 14, no. 9, 2022, Art. no. 2246.

[283] Y. Xu, Y. Zhang, T. Yue, C. Yu, and H. Li, "Graph-based domain adaptation few-shot learning for hyperspectral image classification," *Remote Sens.*, vol. 15, no. 4, 2023, Art. no. 1125.

[284] H. Wang, X. Wang, and Y. Cheng, "Graph meta transfer network for heterogeneous few-shot hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5501112.

[285] Q. Ran, "Deep transformer and few-shot learning for hyperspectral image classification," *CAAI Trans. Intell. Technol.*, vol. 8, pp. 1323–1336, 2023.

[286] Y. Peng, Y. Liu, B. Tu, and Y. Zhang, "Convolutional transformer-based few-shot learning for cross-domain hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 1335–1349, 2023.

[287] Z. Huang, H. Tang, Y. Li, and W. Xie, "HFC-SST: Improved spatial-spectral transformer for hyperspectral few-shot classification," *J. Appl. Remote Sens.*, vol. 17, no. 2, 2023, Art. no. 026509.

[288] A. Creswell, T. White, V. Dumoulin, K. Arulkumaran, B. Sengupta, and A. A. Bharath, "Generative adversarial networks: An overview," *IEEE Signal Process. Mag.*, vol. 35, no. 1, pp. 53–65, Jan. 2018.

[289] 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.

[290] 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, 2017, Art. no. 1042.

[291] 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.

[292] 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.

[293] 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.

[294] 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.

[295] 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.

[296] Z. Xue, "Semi-supervised convolutional generative adversarial network for hyperspectral image classification," *IET Image Process.*, vol. 14, no. 4, pp. 709–719, 2020.

[297] 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.

[298] 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.

[299] 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, 2021, Art. no. 5500615.

[300] D. Hong, "Cross-city matters: A multimodal remote sensing benchmark dataset for cross-city semantic segmentation using high-resolution domain adaptation networks," *Remote Sens. Environ.*, vol. 299, 2023, Art. no. 113856.

[301] 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.

[302] 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.

[303] 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, 2018, Art. no. 1425.

[304] O. Day and T. M. Khoshgoftaar, "A survey on heterogeneous transfer learning," *J. Big Data*, vol. 4, pp. 1–42, 2017.

[305] 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.

[306] 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.

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

[308] 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.

[309] R. P. De Lima and K. Marfurt, "Convolutional neural network for remote-sensing scene classification: Transfer learning analysis," *Remote Sens.*, vol. 12, no. 1, 2019, Art. no. 86.

[310] D. Hong et al., "SpectralGPT: Spectral foundation model," 2023, *arXiv:2311.07113*.

[311] X. He, Y. Chen, and Z. Lin, "Spatial-spectral transformer for hyperspectral image classification," *Remote Sens.*, vol. 13, no. 3, 2021, Art. no. 498.

[312] Y. Qing, W. Liu, L. Feng, and W. Gao, "Improved transformer net for hyperspectral image classification," *Remote Sens.*, vol. 13, no. 11, 2021, Art. no. 2216.

[313] L. Sun, G. Zhao, Y. Zheng, and Z. Wu, "Spectral–spatial feature tokenization transformer for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5522214.

[314] B. Liu, A. Yu, K. Gao, X. Tan, Y. Sun, and X. Yu, "DSS-TRM: Deep spatial–spectral transformer for hyperspectral image classification," *Eur. J. Remote Sens.*, vol. 55, no. 1, pp. 103–114, 2022.

[315] H. Yu, Z. Xu, K. Zheng, D. Hong, H. Yang, and M. Song, "MSTNET: A multilevel spectral–spatial transformer network for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5532513.

[316] L. Yang, "FusionNet: A convolution–transformer fusion network for hyperspectral image classification," *Remote Sens.*, vol. 14, no. 16, 2022, Art. no. 4066.

[317] J. Zhang, Z. Meng, F. Zhao, H. Liu, and Z. Chang, "Convolution transformer mixer for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2022, Art. no. 6014205.

[318] C. Zhao, "Hyperspectral image classification with multi-attention transformer and adaptive superpixel segmentation-based active learning," *IEEE Trans. Image Process.*, vol. 32, pp. 3606–3621, 2023.

[319] Z. Qiu, J. Xu, J. Peng, and W. Sun, "Cross-channel dynamic spatial-spectral fusion transformer for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5528112.

[320] S. K. Roy, A. Deria, C. Shah, J. M. Haut, Q. Du, and A. Plaza, "Spectral–spatial morphological attention transformer for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5503615.

[321] S. Hao, Y. Xia, and Y. Ye, "Generative adversarial network with transformer for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 20, 2023, Art. no. 5510205.

[322] X. Cao, H. Lin, S. Guo, T. Xiong, and L. Jiao, "Transformer-based masked autoencoder with contrastive loss for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5524312.

[323] L. Liang, Y. Zhang, S. Zhang, J. Li, A. Plaza, and X. Kang, "Fast hyperspectral image classification combining transformers and SimAM-based CNNs," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5522219.

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

[325] P. Melville and R. J. Mooney, "Diverse ensembles for active learning," in *Proc. 21st Int. Conf. Mach. Learn.*, 2004, p. 74.

[326] C. C. Aggarwal, X. Kong, Q. Gu, J. Han, and S. Y. Philip, "Active learning: A survey," in *Data Classification*. London, U.K.: Chapman & Hall/CRC Press, 2014, pp. 599–634.

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

[328] B. Settles, "Active learning literature survey," Univ. Wisconsin—Madison, Madison, WI, USA, Tech. Rep. 1648, 2009.

[329] 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. 2018.

[330] 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, 2019, Art. no. 2974.

[331] 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.

[332] 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.

[333] 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. Proc.*, 2020, pp. 1219–1229.

[334] 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.

[335] 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.

[336] 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.

[337] 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.

[338] 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.

[339] 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.

[340] 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.

[341] 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.

[342] 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.

[343] 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.

[344] 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. 2019.

[345] 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.

[346] M. Ahmad, "Multiclass non-randomized spectral–spatial active learning for hyperspectral image classification," *Appl. Sci.*, vol. 10, no. 14, 2020, Art. no. 4739.

[347] J. Yao et al., "Semi-active convolutional neural networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5537915.

[348] R. E. Schapire, "The boosting approach to machine learning: An overview," in *Nonlinear Estimation and Classification*. New York, NY, USA: Springer, 2003, pp. 149–171.

[349] R. E. Schapire, "Explaining adaboost," in *Empirical Inference: Festschrift Honor Vladimir N. Vapnik*. Berlin, Germany: Springer, 2013, pp. 37–52.

[350] G. Biau and E. Scornet, "A random forest guided tour," *Test*, vol. 25, pp. 197–227, 2016.

[351] B. Waske, S. Van Der Linden, J. A. Benediktsson, A. Rabe, and P. Hostert, "Sensitivity of support vector machines to random feature selection in classification of hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 7, pp. 2880–2889, Jul. 2010.

[352] W. Yao, C. Lian, and L. Bruzzone, "A CNN ensemble based on a spectral feature refining module for hyperspectral image classification," *Remote Sens.*, vol. 14, no. 19, 2022, Art. no. 4982.

[353] J. Nalepa, M. Myller, L. Tulczyjew, and M. Kawulok, "Deep ensembles for hyperspectral image data classification and unmixing," *Remote Sens.*, vol. 13, no. 20, 2021, Art. no. 4133.

[354] S. Dong, W. Feng, Y. Quan, G. Dauphin, L. Gao, and M. Xing, "Deep ensemble CNN method based on sample expansion for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5531815.

[355] Y. Chen, Y. Wang, Y. Gu, X. He, P. Ghamisi, and X. Jia, "Deep learning ensemble for hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 6, pp. 1882–1897, Jun. 2019.

[356] H. Su, Y. Hu, H. Lu, W. Sun, and Q. Du, "Diversity-driven multikernel collaborative representation ensemble for hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 2861–2876, 2022.

[357] H. Lu, H. Su, J. Hu, and Q. Du, "Dynamic ensemble learning with multi-view kernel collaborative subspace clustering for hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 2681–2695, 2022.

[358] Q. Lv, W. Feng, Y. Quan, G. Dauphin, L. Gao, and M. Xing, "Enhanced-random-feature-subspace-based ensemble CNN for the imbalanced hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 3988–3999, 2021.

[359] Q. Lv, "Ensemble CNN with enhanced feature subspaces for imbalanced hyperspectral image classification," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2021, pp. 3669–3672.

[360] Q. Li, B. Zheng, B. Tu, J. Wang, and C. Zhou, "Ensemble EMD-based spectral-spatial feature extraction for hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 5134–5148, 2020.

[361] H. Su, Y. Yu, Q. Du, and P. Du, "Ensemble learning for hyperspectral image classification using tangent collaborative representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 6, pp. 3778–3790, Jun. 2020.

[362] W. Feng, "Ensemble margin based semi-supervised random forest for the classification of hyperspectral image with limited training data," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2019, pp. 971–974.

[363] B. Liu, K. Gao, A. Yu, L. Ding, C. Qiu, and J. Li, "ES2FL: Ensemble self-supervised feature learning for small sample classification of hyperspectral images," *Remote Sens.*, vol. 14, no. 17, 2022, Art. no. 4236.

[364] R. Agarwal, A. Aziz, A. S. Krishnan, A. Challa, and S. Danda, "ESW edge weights: Ensemble stochastic watershed edge weights for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2022, Art. no. 6009605.

[365] X. Yin, R. Wang, X. Liu, and Y. Cai, "Deep forest-based classification of hyperspectral images," in *Proc. 37th Chin. Control Conf.*, 2018, pp. 10367–10372.
[366] T. Li, J. Leng, L. Kong, S. Guo, G. Bai, and K. Wang, "DCNR: Deep cube CNN with random forest for hyperspectral image classification," *Multimedia Tools Appl.*, vol. 78, pp. 3411–3433, 2019.
[367] X. Liu, R. Wang, Z. Cai, Y. Cai, and X. Yin, "Deep multigrained cascade forest for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 10, pp. 8169–8183, Oct. 2019.
[368] X. Cao, L. Wen, Y. Ge, J. Zhao, and L. Jiao, "Rotation-based deep forest for hyperspectral imagery classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 7, pp. 1105–1109, Jul. 2019.
[369] X. Cao, R. Li, Y. Ge, B. Wu, and L. Jiao, "Densely connected deep random forest for hyperspectral imagery classification," *Int. J. Remote Sens.*, vol. 40, no. 9, pp. 3606–3622, 2019.
[370] B. Liu, "Morphological attribute profile cube and deep random forest for small sample classification of hyperspectral image," *IEEE Access*, vol. 8, pp. 117096–117108, 2020.
[371] J. Shi, T. Shao, X. Liu, X. Zhang, Z. Zhang, and Y. Lei, "Evolutionary multitask ensemble learning model for hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 936–950, 2021.
[372] X. Liu, Q. Hu, Y. Cai, and Z. Cai, "Extreme learning machine-based ensemble transfer learning for hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 3892–3902, 2020.
[373] W. Feng, W. Huang, and W. Bao, "Imbalanced hyperspectral image classification with an adaptive ensemble method based on SMOTE and rotation forest with differentiated sampling rates," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 12, pp. 1879–1883, Dec. 2019.
[374] Y. Zhang, G. Cao, A. Shafique, and P. Fu, "Label propagation ensemble for hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 9, pp. 3623–3636, Sep. 2019.
[375] L. Fang, W. Zhao, N. He, and J. Zhu, "Multiscale CNNs ensemble based self-learning for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 9, pp. 1593–1597, Sep. 2020..
[376] Y. Zhang, G. Cao, and X. Li, "Multiview-based random rotation ensemble pruning for hyperspectral image classification," *IEEE Trans. Instrum. Meas.*, vol. 70, 2021, Art. no. 5500314.
[377] W. Feng, "Semi-supervised rotation forest based on ensemble margin theory for the classification of hyperspectral image with limited training data," *Inf. Sci.*, vol. 575, pp. 611–638, 2021.
[378] X. He and Y. Chen, "Transferring CNN ensemble for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 18, no. 5, pp. 876–880, May 2021.
[379] W. Feng, "Two-step ensemble based class noise cleaning method for hyperspectral image classification," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2020, pp. 830–833.
[380] H. Lu, H. Su, P. Zheng, Y. Gao, and Q. Du, "Weighted residual dynamic ensemble learning for hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 6912–6927, 2022.
[381] J. Yao, B. Zhang, C. Li, D. Hong, and J. Chanussot, "Extended vision transformer (ExViT) for land use and land cover classification: A multimodal deep learning framework," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5514415.
[382] J. Yao, D. Hong, H. Wang, H. Liu, and J. Chanussot, "UCSL: Towards unsupervised common subspace learning for cross-modal image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5514212.
[383] D. Hong, J. Yao, D. Meng, Z. Xu, and J. Chanussot, "Multimodal GANs: Toward crossmodal hyperspectral–multispectral image segmentation," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 6, pp. 5103–5113, Jun. 2021.
[384] H. Zhang, J. Yao, L. Ni, L. Gao, and M. Huang, "Multimodal attention-aware convolutional neural networks for classification of hyperspectral and LiDAR data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 3635–3644, 2023.
[385] 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.
[386] 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.
[387] 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.
[388] X. Xu, J. Li, and A. Plaza, "Kennedy space center dataset," Accessed: Jan. 30, 2023. [Online]. Available: https://www.csr.utexas.edu/projects/rs/hrs/classify.html
[389] Z. Ge, G. Cao, X. Li, and P. Fu, "Hyperspectral image classification method based on 2D–3D CNN and multibranch feature fusion," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 5776–5788, 2020.
[390] C. Yu, R. Han, M. Song, C. Liu, and C.-I. Chang, "Feedback attention-based dense CNN for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5501916.
[391] L. Mou, X. Lu, X. Li, and X. X. Zhu, "Nonlocal graph convolutional networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 12, pp. 8246–8257, Dec. 2020.
[392] F. Ullah, "Deep hyperspectral shots: Deep snap smooth wavelet convolutional neural network shots ensemble for hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 17, pp. 14–34, 2024.
[393] Y. Wang et al., "Heterogeneous few-shot learning for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.

Farhan Ullah received the B.Sc. degree from the University of Peshawar, Peshawar, Pakistan, in 2013, and the M.C.S. degree from the Sarhad University of Science and Information Technology, Peshawar, in 2015, both in computer science, and the Ph.D. degree in computer science and technology from Shanghai University, Shanghai, China, in 2020.

He is currently a Postdoctoral Fellow with the College of Computer Science and Software Engineering, Shenzhen University, Shenzhen, China. His research interests include deep learning, hyperspectral image classification, and computer vision.

Irfan Ullah received the B.Sc. and M.Sc. degrees in electrical engineering from the Sarhad University of Science and Information Technology, Peshawar, Pakistan, in 2012 and 2015, respectively, and the Ph.D. degree in electrical engineering from the State Key Laboratory of Power Transmission Equipment and System Security and New Technology, School of Electrical Engineering, Chongqing University, Chongqing, China, in 2019.

In 2021, he was a Postdoctoral Researcher with the School of Engineering, Chulalongkorn University, Bangkok, Thailand. He is currently an Associate Professor with the School of Computer Science, Chengdu University of Technology, Chengdu, China. He has authored or coauthored several papers in conferences and journals, such as IEEE TRANSACTIONS ON POWER ELECTRONICS, IEEE TRANSACTIONS ON MAGNETICS, *Applied Thermal Engineering*, and IEEE ACCESS.

Rehan Ullah Khan received the B.Sc. and M.Sc. degrees in information systems from the University of Engineering and Technology, Peshawar, Pakistan, in 2004 and 2006, respectively, and the Ph.D. degree in artificial intelligence from the Vienna University of Technology, Vienna, Austria, in 2011.

He is currently an Associate Professor with the Department of Information Technology, College of Computer, Qassim University, Buraydah, Saudi Arabia. His current research interests include machine learning, segmentation, and image-based object recognition.

Salabat Khan received the M.Sc. degree in computer science from the University of Peshawar, Peshawar, Pakistan, and the M.S. degree in computer science from the International Islamic University, Islamabad, Pakistan, both in computer science, and the Ph.D. degree in computer science and technology from the Beijing Institute of Technology, Beijing, China.

He is currently a Postdoctoral Fellow with the College of Computer Science and Software Engineering, Shenzhen University, Shenzhen, China. His main research interests include machine learning, deep learning, security and privacy issues on the Internet, Internet of Things, and Internet of Vehicles.

Khalil Khan received the B.S. degree in electrical engineering and the M.S. degree in computer engineering from the University of Engineering and Technology, Peshawar, Pakistan, in 2007 and 2012, respectively, and the Ph.D. degree in intellegence and machine learning from the Signals and Communication Laboratory, University of Brescia, Brescia, Italy, in 2016.

He is currently an Assistant Professor with the Department of Computer Science, School of Engineering and Digital Sciences, Nazarbayev University, Astana, Kazakhstan.

Giovanni Pau (Member, IEEE) received the bachelor's degree in telematic engineering from the University of Catania, Catania, Italy, in 2008, and the master's degree (cum laude) in telematic engineering and the Ph.D. degree from Kore University of Enna, Enna, Italy, in 2010 and 2015.

He is currently an Associate Professor with the Faculty of Engineering and Architecture, Kore University of Enna. He is the author/coauthor of more than 80 refereed articles published in journals and conferences proceedings. He has been involved in the organization of several international conferences as a session cochair and technical program committee member. He serves/served as a leading Guest Editor in special issues of several international journals. His research interests include wireless sensor networks, fuzzy logic controllers, intelligent transportation systems, Internet of Things, and network security.

Dr. Pau is an Editorial Board Member as an Associate Editor for IEEE ACCESS, *Wireless Networks*, *EURASIP Journal on Wireless Communications and Networking*, *Wireless Communications and Mobile Computing*, and *Future Internet*.