

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

Farhan Ullah, Irfan Ullah, Rehan Ullah Khan, Salabat Khan, Khalil Khan, Giovanni Pau

Abstract—*Hyperspectral image classification has become a hot research topic. 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 Hyperspectral Image Classification (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 the 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—deep ensemble, hyperspectral image classification, CNN, deep learning, spectral features, spatial 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]. Hyperspectral imaging provides rich spectral information by acquiring reflectance measurements across hundreds of narrow, contiguous spectral bands over a wide wavelength range from $0.4 - 2.4\mu m$. This span covers the

visible region from $0.4 - 0.7\mu m$ as well as the shortwave infrared region from $0.7 - 2.4\mu m$. 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 facilitates 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 investigate the (light) emission features of instances in the mid to long-infrared range [3].

Alongside the comprehensive information, it poses some issues considering conventional analysis methods for monochromatic, RGB, and multispectral images 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 the atmosphere, climate, urban, agricultural production, geomorphological and mineral investigation, coastal area, maritime, forest management, groundwater and surface contamination, waterways, glaciers and ice sheets, biological, healthcare cases, and food computation [4]–[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]–[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 is 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 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. Figure 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

Corresponding author: Giovanni Pau

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

Irfan Ullah is with the School of Computer Science, Chengdu University of Technology, Sichuan, China (irfan.ee@cqu.edu.cn)

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

Khalil Khan is with the Department of computer science, Nazarbayev university Kazakhstan (khalil.khan@nu.edu.kz)

Giovanni Pau is with the Kore University of Enna, Italy (giovanni.pau@unikore.it)

this methodology.

Given the limitations mentioned above, HSI evaluation is divided into the following categories: Dimensionality Reduction (DR) [14]–[18] have enabled more efficient processing of hyperspectral data by extracting the most meaningful features and reducing redundant information. Spectral unmixing [19]–[27] have 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]–[34] have utilized spectral-spatial approaches to identify areas of interest and track their evolution over time. Classification techniques [35]–[37] incorporating spatial context have improved mapping accuracy compared to pixel-wise classifiers, feature learning for classification [38]–[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. Figure 1b 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]–[51], environmental surveillance, and hazard identification [52], [53] vegetation mapping [54], [55], and urban planning. To complete the classification task, HSIC techniques use machine learning techniques [56]–[58]. These techniques are described in numerous in-depth reviews that were published over the past decade [35], [59]–[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]–[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 Deep Learning with Ensemble (DLE) 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:

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

where d denotes the total number of spectral bands, with each band containing $(P \times Q)$ samples per band belonging to Z classes, where $c_i = [c_{1_i}, c_{2_i}, c_{3_i}, \dots, c_{n_i}]T$ is the i th sample in the HSI cube with class label $y_i \in R^y$. Hyperspectral classification aims to learn a mapping function $f_v(\cdot)$ 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 $f_v(\cdot)$ maps the high-dimensional input data to the label space by exploiting the rich spectral information and inherent class structure to achieve accurate classification.

$$Z = f_v(U, \theta) \quad (2)$$

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

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 [73], Histogram of Oriented Gradients (HOG) [74], Global Image Scale-Invariant Transform/Global Invariant Scalable Transform (GIST) [75], pyramid HOGs, 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 feature extraction and discriminative classification has been shown to enhance mapping accuracy compared to pixel-wise 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

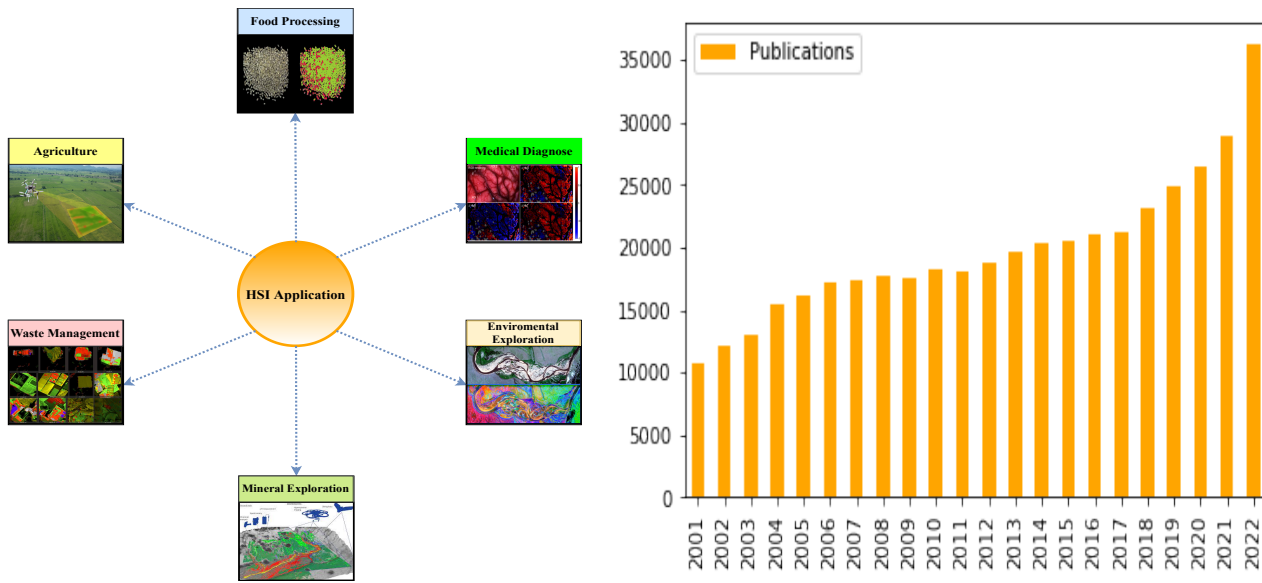


Fig. 1 (a) Various real-world applications of HSI and (b) Remote sensing/HSI-relevant research papers published per year till April 25, 2023 [Source: Google Scholar accessed on April 25, 2023, and the total number of research papers (including patents and citations) were classified by its relevancy].

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]–[81].

SIFT is a widely applied vigorous feature descriptor used for computer vision applications [82]–[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 sub-regions 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 local binary patterns (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]–[95], and HOG feature-based models [47], [96]. 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]–[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]–[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. Additionally, 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, the authors in [113] introduced a novel baseline network for HAD, named LRR-Net, which combines the LRR model with deep learning methods. LRR-Net efficiently employs the alternating direction method of multipliers (ADMM) 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 non-linear features from the input data without any pre-existing 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, semi-supervised, 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. Additionally, 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]. Below are some major issues that arise when DL is performed on HSIC.

- 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]–[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 (GC) [129], and AngularGrad [130], respectively, are effective CNN optimization approaches that are commonly employed in classification tasks.
- 2) **Training Data Scarcity Issue:** As previously stated, supervised DNN needs a significant size of training samples; otherwise, its inclination to over-fit 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 amounts of 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 DNN retrieves increasingly rich information from data [135]; Although, 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 is 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 bands and N and M represent spatial features, i.e., width and height, respectively. The HSI cube is exhibited in Figure 2.

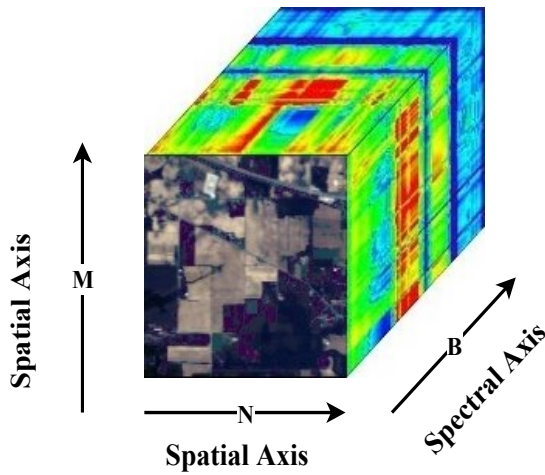


Fig. 2 Illustrates hyperspectral spatial and spectral cube

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. Dimensionality reduction 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 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 dimensionality reduction 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. Although, 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 hyperspectral imaging. 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), Local Binary Patterns (LBP), etc., and DNN-based algorithms are some popular techniques for extracting spatial information from HSI cubes. Geometrical properties can be extracted from morphological profiles. Some improvement in MPs involves Extended Morphological Profiles (EMPs) [142], multiple-structure-element morphological profiles [143], and invariant attribute profiles [144].

The image processing technique produces valuable spatial contextual features for HSI. For example, a texture analysis tool called a Gabor filter 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. Additionally, the spatial features included in HSI can also be retrieved by merging previously mentioned approaches. For example, [145] merged the GF method and Differential Morphological Profiles (DMPs) method [146] to retrieve Local Spatial Sequential (LSS) features for an 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 deeply execute the 3-D HSI cube 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]–[156]. In section 5, 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 hyper-parameters 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 hyper-parameters [157]–[159]. 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. Although, such learning algorithms are effective when we aim to extract the intrinsic pattern of such datasets with limited training samples. The 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. Semi-Supervised Learning Approach

Semi-supervised learning (SSL) is the third approach that falls between unsupervised and supervised methodologies. SSL techniques learn from partly labeled dataset, i.e., a limited number 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 [160]–[163]. In Section 11, the specifics of these techniques are briefly discussed.

VI. DNNs CONSTRUCTION (LAYERS 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 non-linearly, 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 below.

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 hyper-parameters that need to be tweaked. Additionally, the model is highly prone to overfitting problems due to the significant number of hyper-parameters [50]. Therefore, a dropout approach is adopted in order to lessen the impact of the overfitting issue [164].

B. Convolutional Layers

In HSIC, convolutional layers are a fundamental component of Convolutional Neural Networks (CNNs) designed to process hyperspectral data [165]. 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 hyperspectral image classification is given below:

1) Input Representation:

- **3D Tensor:** Hyperspectral images are represented as 3D 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:

- **Spatial Convolution:** Traditional 2D convolutional operations are applied to capture spatial patterns and relationships between pixels. Filters are used to detect features in the spatial domain.
- **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.

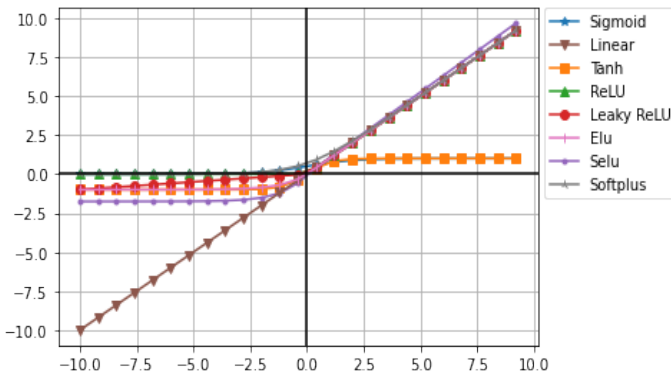


Fig. 3 Shows a graphical depiction of different frequently employed activation functions.

3) *Feature Extraction:*

- **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:*

- **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:*

- **Shared Weights:** Convolutional layers use shared weights across the input space, reducing the number of parameters compared to fully connected layers. This parameter sharing enhances the model’s ability to generalize to different spatial locations and spectral bands.

C. *Activation Layers*

In Deep Neural Networks (DNNs), the activation layers are considered the feature detector phase [166] because they play a crucial role in capturing and learning relevant patterns and representations from the input data. The FC and CONV layers either produce linear representations of input data or treat them in a manner analogous to linear regressors, and the data that undergoes transformation by these layers is described as being in the FE phase [69]. As a result, an activation layer must be utilized after the FC layer and Conv 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 [167], and SoftMax are a few examples of regularly applied activation functions. Although, the activation functions SoftMax and ReLU are frequently used in HSI analysis [71]. A few frequently used activation functions are shown graphically in Figure 3.

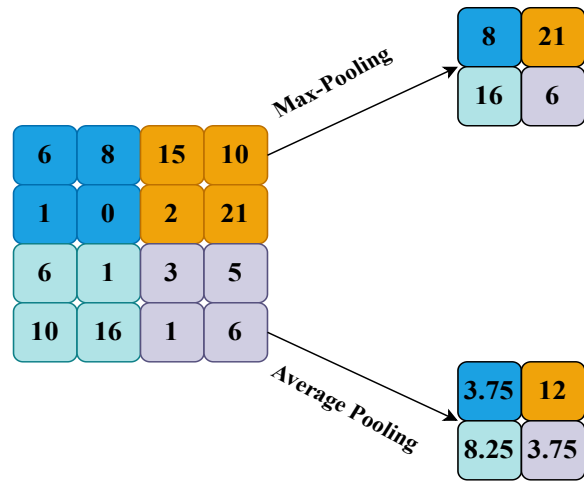


Fig. 4 The graphical depictions of the Max-pooling and average-pooling procedures of the downsampling/pooling layer.

D. *Pooling Layers*

The pooling layers are also referred to as the down-sampling layer, which accepts a given input data size and transforms it into a single value, as illustrated in Figure 4. This enables in-variance to minimal data anomalies. Since the volume of the input and the model hyper-parameters are both minimized, which also reduces the computing time, the pooling layer aids the model in preventing over-fitting problems. Among the various down-sampling techniques, the three most popular ones include max-pooling, average-pooling, and sum-pooling. Furthermore, a new optimized pooling method called wavelet-pooling was presented in [168], and its effectiveness is comparable to that of max-pooling and average pooling. Similarly, [169] presented an alternative approach in which the convolutional layer of enhanced filter stride takes the role of the pooling layer.

VII. CONVOLUTIONAL NEURAL NETWORK

The Convolutional Neural Network (CNN) architecture draws inspiration from the biological visual system, as described in [170]. Adopting Hubel and Wiesel’s natural visual recognition process [170], Neocognitron [171] is recognized as the original hierarchical, position-invariant pattern recognition model [172], and can be regarded as the prior of CNN [173]. 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 convolution 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 hyper-parameters 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 hyper-parameters required to be tweaked by the

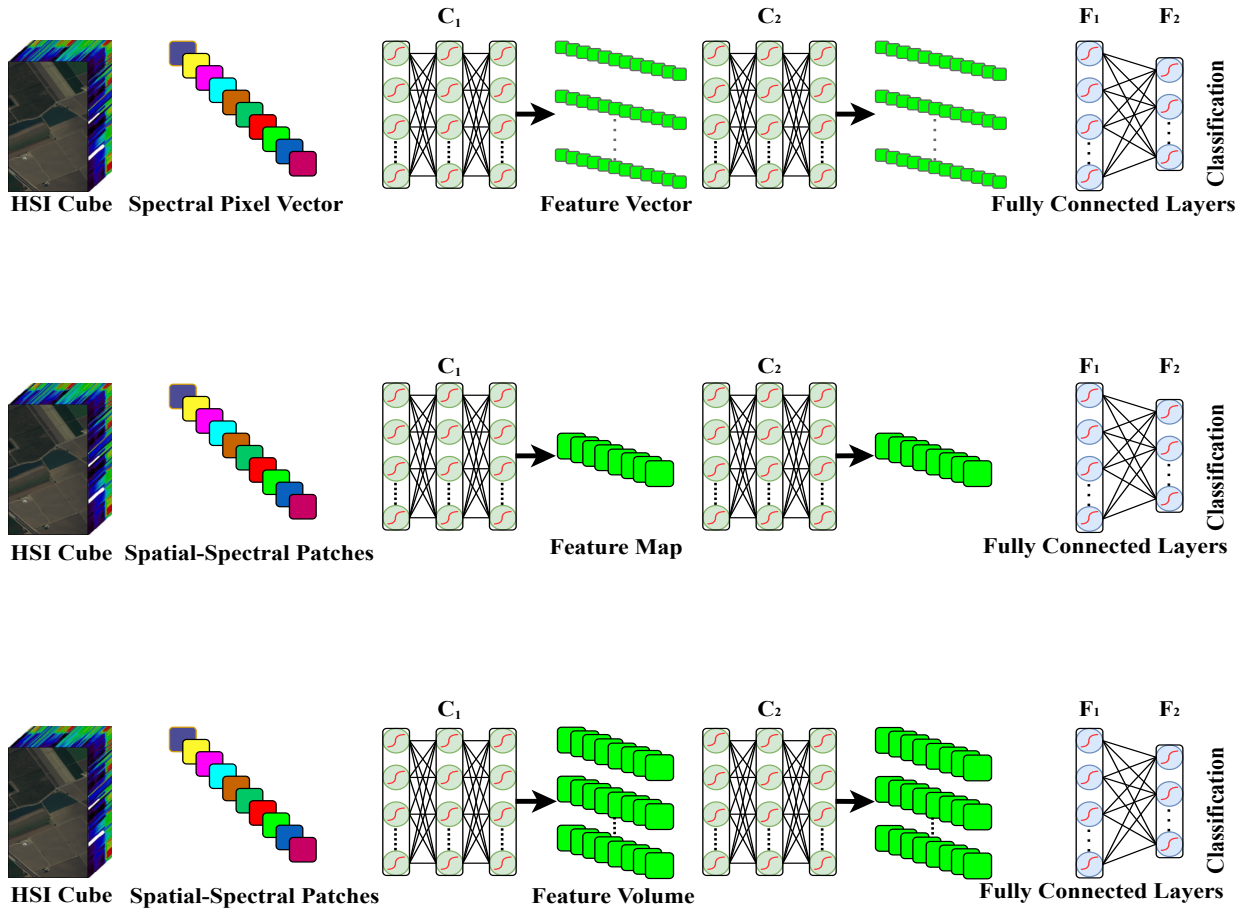


Fig. 5 Shows the generic structure of spectral CNN, spatial CNN, and spectral-spatial CNN frameworks for HSIC.

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 non-linearities to the network. This is accomplished by taking the outputs of 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 author presented in [174], [175] the FC layer can be skipped by employing a global average pooling layer. SoftMax is frequently employed for classification problems [176], [177], although SVM [178], [179] 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: i) Spectral CNN frameworks, ii) spatial CNN frameworks, and iii) spectral-spatial frameworks CNN. Figure 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 bands retrieved after the DR method. In the research study [180], a CNN architecture was presented to reduce the over-fitting issue and acquire improved generalization ability via using the 1×1 convolutional kernels and increased dropout rates. Furthermore, the incorporation of a Global Average Pooling (GAP) layer in lieu of a fully connected (FC) layer serves a dual purpose in the neural network architecture. It not only effectively reduces the overall number of network hyper-parameters, but it also brings about a notable enhancement in the model’s ability to generalize and avoid overfitting. In a study [175], 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 Convolution layer. In [175], the authors used the (GAP) layer in a similar way [180] 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 by Wu and Prasad [181], 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 [182], it was shown that 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 [183], PCA was used to preprocess medical HSI before CNN kernels and Gabor kernels were fused with dot products to perform classification.

In [184], the authors 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 [185], a technique of averaged spectrum is employed to minimize the noise impact in pixel-wise HSIC, where an averaged spectrum of a set of pixels corresponding to bacterial colonies is 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. [186] 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. [187] used a single PC to train a spatially-based 2D-CNN. In [188], the input data preprocessed using PCA whitening and limited to three principal components (PCs), is fed into a randomized patches network, which serves as a 2D-CNN classification architecture. Although, a small number of training instances with relatively comparable spectral features cause DL models to be sensitive to the overfitting problem. In order to address this, Wang et al. [189] presented a novel approach called Probabilistic Neighborhood Pooling-based Attention Network (PNPAN) specifically designed for HSIC.

The approach presented in study [190], exploited patches clipped from 2-D input instances (i.e., images from various spectral bands) to train a 2D-CNN framework that acquires the data-adaptive kernels on its own. Moreover, several authors presented using handcrafted-based features in conjunction with spectral-domain minimization. For instance, to avoid the overfitting issue caused by insufficient training data, [191] integrated the Gabor-filtering method with 2D-CNN for HSIC. Gabor filtering captures spatial features such as edges and textures, significantly reducing the overfitting issue. The study in [192], 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. [193] introduced a better Pixel Pair Feature (PPF) method referred to as spatial PPF. This method differs from conventional PPFs in two key ways: firstly, 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 secondly, the label of the pixel pair would be as of the central pixel. In another study, [194], the authors 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 3D-CNN approach that effectively uses both spectral and spatial data for HSIC was put forth in [195].

Recent work has developed advanced deep-learning techniques to extract informative features from hyperspectral images while modeling spatial context. Li et al. [196] 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 feature extraction. This creates more versatile convolutional kernels. As studied in [197], combining max and min convolutional features before the ReLU activation can further improve the discriminative capability of extracted features for classification. Paoletti et al. [198] 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, DNN networks 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 the article [199] to solve the problem of the HSIC's scarcity of training instances. Jia et al. [200] introduced a novel approach to extracting the spatial-spectral features by Schroedinger Eigen maps 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, the paper [201] 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, the paper [202] proposed a novel morphological framework called end-to-end morphological CNN (MorphCNN) for HSIC, integrating both spectral and spatial information through hybridizing the outputs of spectral and spatial morphological blocks retrieved in a dual-path framework.

The paper [196] presented a duplet-phase approach for joint spectral-spatial HSIC that can effectively generate spectral and spatial features rather than hybridizing them separately. The presented approach's first phase 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 (SAE), 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 3D-CNN model was presented in [203]. To analyze its effectiveness, a comparative analysis is made with spectral-based DBN, SAE, and 2-D spatial CNN for HSIC. The study [204] presented a novel Fused squeeze-and-excitation network (Fusenet), which is a bi-linear 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 [205]. 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.

In [206], the authors proposed an efficient and novel dual-channel CNN architecture for spectral-spatial HSIC tasks. The presented method first employed a 1D-CNN to retrieve hierarchical spectral information, and 2D-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 [207] 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. In [208], the authors 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 [209], 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 [210], the proposed method merged CNN with the metric learning-based HSIC approach. In the first phase, they use CNN 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. Additionally, [211] combines a multiscale convolution-based CNN architecture with diverse deep metric learning techniques based on Determinantal Point Process (DPP) priors [212]. This framework extracts multiscale features across 1D spectral, 2D spectral-spatial, and 3D spectral-spatial modalities. 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 feature extraction 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.

In [213], the authors 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.

In [214], they 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. The authors in [215], presented a novel adaptive band selection-based semi-supervised 3-D-CNN approach for jointly exploiting spectral information, whereas [216] investigated a dual-attention-based autoencoder-decoder DNN for unsupervised hyperspectral band selection and consequently merge FE for land-cover classification. Correspondingly, spectral-spatial features are utilized in an unsupervised approach in [217] utilizing a 3-D convolution AE. The inclusion of incorrect/noisy labels in the training sample frequently causes pixel-wise land-use and land-cover classification employing conventional CNNs to be overfitted to the labeled noises. To address the issue of efficient classification, Roy et al. [218] 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. The authors in [219], introduce 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. [220] proposed a hybrid HybridSN framework in which joint spectral-spatial features are initially retrieved using 3D-CNN, and following the initial feature extraction, the 2D-CNN plays a crucial role in further refining the representation of the data by retrieving extra abstract spatial contextual information. The finding in [221], presented employing adaptive Markov random field for HSIC. Once the CNN has effectively captured joint spectral-spatial features, the next step involves applying smooth Markov Random Field (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. [222] 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, [223] 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 convolutions process. G2C-Conv3D can be effectively incorporated into the current HSIs FE networks to enhance the effectiveness of effective land types classification. Recently, the authors in [224], introduce a novel method, Hashing-based Deep Metric Learning (HDML), designed for classifying hyperspectral images (HSIs) 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 Convolutional Neural Network Frameworks for HSIC

Graph Convolutional Neural Networks (GCNs) [225] 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 non-grid high-dimensional data. Such characteristics offer new insights and opportunities for more effective and efficient computation of hyperspectral data. GCNs, in particular, allow for the modeling of data relationships. As a result, we are generally motivated to employ GCNs to acquire the spatial relationships of spectral signatures in HSIs. Despite GCNs' constraints in graph construction [226], especially for sizeable graphs (requiring costly computational complexity), GCNs 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 GCNs in the HSIC function.

For instance, [227] a second-order Graph Convolutional Network (SO-GCN), a novel approach to modeling spatial

features on manifolds for Hyperspectral Image Classification (HSIC). The primary objective of this innovation was to reduce the computational burden associated with graph processing. Adding to this discourse, Wan et al. [228] pioneered the super-pixel segmentation method on HSIs, feeding super-pixels into GCNs instead of individual pixels. This methodology enabled the training phase of GCNs 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 GCNs. To bridge this gap, Hong et al. [226] 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; thus, 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 is 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 CNN with unsupervised methods. Additionally, 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, GCNs 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

Autoencoders (AEs) are an effective symmetrical neural network for HSIC because of their ability to unsupervised

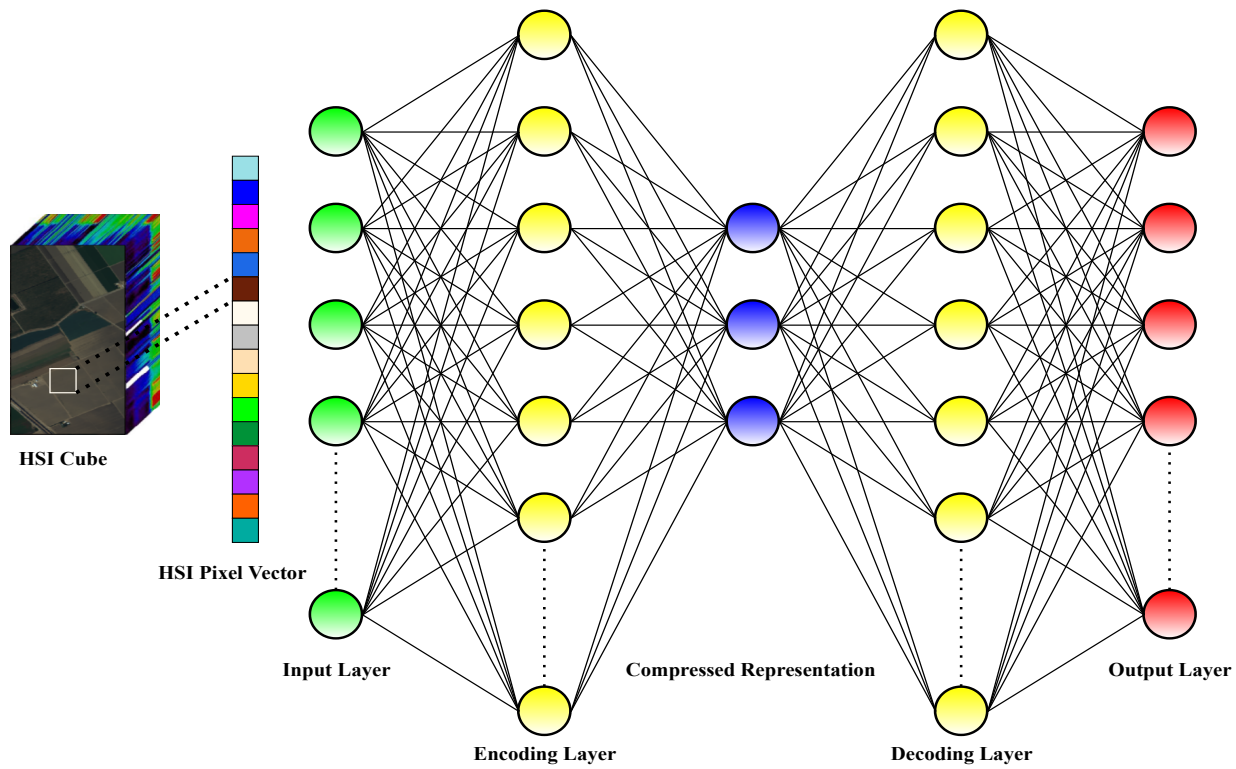


Fig. 6 Shows a generic architecture of autoencoder.

learning. AE provides a compact feature representation of high-dimensional HSI data rather than performing a classification process on its own. As depicted in Figure 6, AE is composed of an input layer, a hidden or encoding layer, a decoding layer, and an output layer. When trained on input data, 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 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, unlike AE, DAE can restore the actual input from a noisy input signal.

In [229], the authors 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 [230] integrated multi-manifold learning [231] approach with a counteractive autoencoder [232] for enhanced unsupervised HSIC. The study [233] utilized spectral-spatial information of HSI using an unsupervised feature-extracting approach comprised of a recursive autoencoder network. It retrieves features from the target pixel's neighborhood and assigns weights according to the spectral correlation between the target and neighboring

pixels. In [234] presented a two DNN with a class-specific fusion framework that learns the fusion weights dynamically. A sequence of stacked DAE is employed in the first stream to retrieve spectral information, and 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 multi-feature-based spectral-spatial HSIC that makes use of sparse AE for high-level FE, guided filters [235] for spatial feature retrieval, and PCA for DR. The approach suggested in [236] 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 [237], 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 autoencoders (AE) for high-level feature extraction to capture essential spectral-spatial information.

In a similar manner, [238] employed stacked sparse AE for spectral-spatial and multifractal representations, as well as other higher-order statistical representations. In [239], an integration of SAE and ELM is presented for HSIC, which partitions the training datasets and infuses them using SAE;

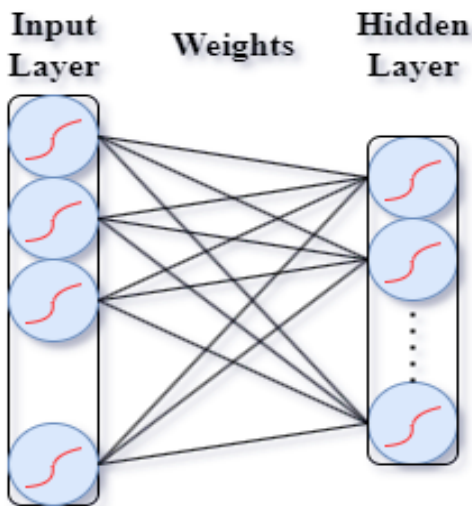


Fig. 7 Shows a basic structure of Restricted Boltzmann Machine.

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. [240] 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. [241] created an SAE-based HSIC that can extract compressed and discriminatory information by introducing a local fisher discriminant regularization. Consequently, the most recent findings [242] combine a k-sparse denoising AE with spectral-restricted spatial features to achieve the high intraclass variance of spatial features for HSIC. The researcher [243] introduced an HSIC framework that initially generates spectral segments of HSI based on joint features criterion to decrease computational complexity during FE via SAE, then incorporates spatial features employing EMPs, and finally uses SVM/RF for classification purposes. Lately, [244] employed 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 AEs-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 inter-class 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 pre-training, co-training, and adaptive neural networks, among others, for AE-based HSIC frameworks could be one key area.

IX. DEEP BELIEF NETWORK

A Deep Belief Network (DBN) [245] 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 Figure 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]. In the article [246] provides a thorough insight into 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. Figure 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 DBN for HSIC. For example, in research study [247], combined spectral-spatial features with DBN for land-cover classification and compared their results to the state-of-the-art classification methods. DBN typically learns in two stages, the first of which is unsupervised pretraining using unlabeled instances and the second of which is supervised fine-tuning using labeled instances. Although, such a training procedure could lead to two issues: First, because of coadaptation [248], multiple hidden units may have a tendency to behave in similar fashion [249], and second, because of the sparseness and specificity of activated neurons, a few of which may always be dead or interacting [250]. In order to address these two issues, Zhong et al. [251] regularized the pretraining and fine-tuning operation by enforcing a diversification before improving the DBN's classification performance for HSI.

In the study mentioned in [252], 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 [253] incorporated a parallel layers 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. [254] 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 [255] presented a spectral-spatial graph-based RBM technique for HSIC, which builds the spectral-spatial graph by measuring combined correlation premised on spectral and

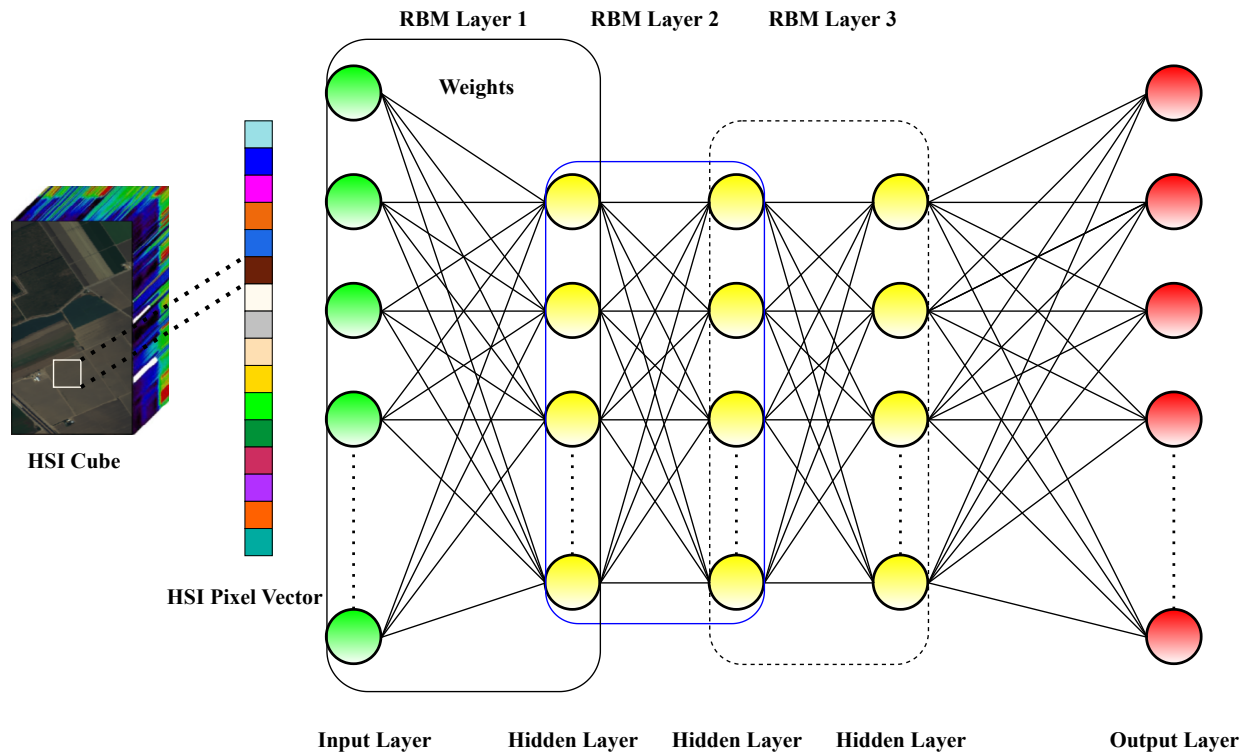


Fig. 8 Shows a three-layer Deep Belief Network architecture.

spatial information, then the model proceeds with training an RBM to extract meaningful joint spectral-spatial features from the 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 semi-supervised learning scenarios, investigating how DBNs can effectively utilize unlabeled or partially labeled hyperspectral data.

Additionally, adapting DBNs to perform robustly with limited datasets is crucial, prompting the exploration of regularization techniques and transfer learning 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 Recurrent Neural Networks (RNNs) structure, depicted in Figure 9, consists of loop connections in which the node activation of the subsequent phase depends on the preceding phase [256]. 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 [257]. RNNs are classified into three types: a) Vanilla, b) Long-Short-Term Memory (LSTM), and c) 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. It learns the related data by substituting the irrelevant data over a period. Although, the gate-controlling approach allows the LSTM a significantly dynamic strategy. The GRU version of LSTM receives the consistency of the Vanilla model while still providing increased accuracy compared to 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. Figure 10 shows a description of the internal architectures of LSTM and GRU.

In the paper [72], the authors introduced a novel RNN-based methodology for HSIC. This approach incorporates a

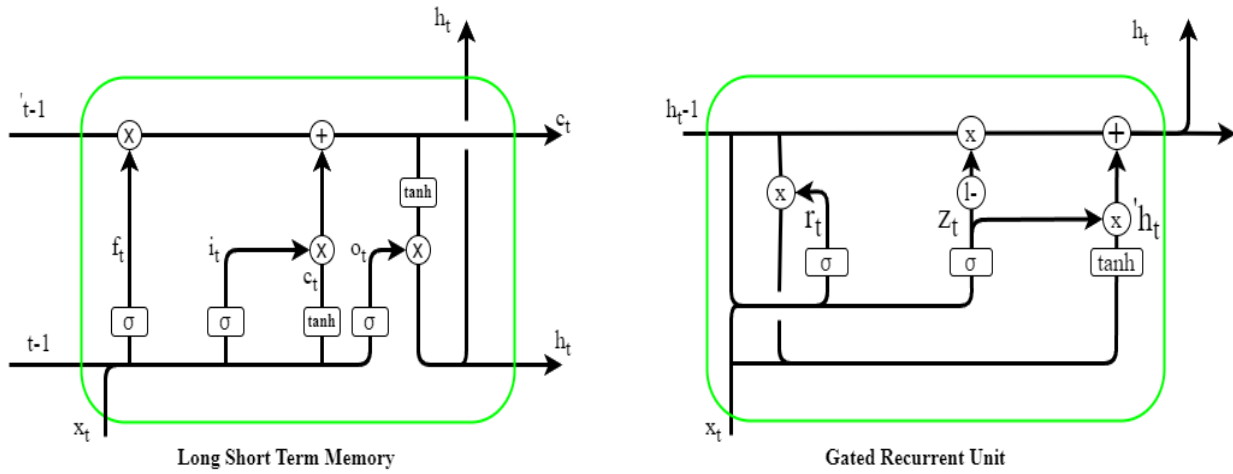


Fig. 9 Shows Internal structure of Long Short Term Memory and Gated Recurrent Unit.

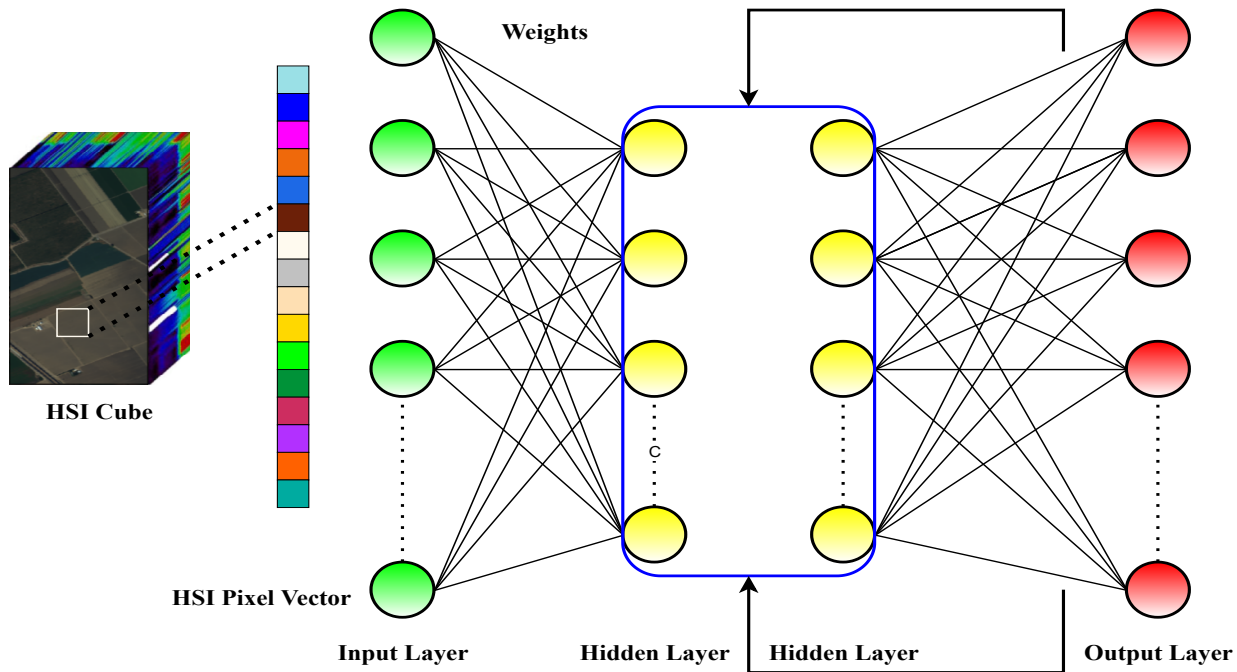


Fig. 10 Shows a Recurrent Neural Network.

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 Gabor filter and differential morphological profiles [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, the study [258] 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. [259] 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 architectures (CRNN) for HSIC. For example, Wu and Prasad [181] 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, in the study [260] used pseudo labels to implement the related model for semi-supervised HSIC. The research article [261] proposed an HSIC approach in which CNN is employed to retrieve spatial information from HSI, and this information is afterward sent to a GRU-based fusion network that conducts feature-level and decision-level fusion.

In a similar way, Luo, et al. [262] leveraged both the spatial and spectral data inherent in HSI by integrating a Convolutional Neural Network (CNN) with a parallel Gated Recurrent Unit (GRU)-based Recurrent Neural Network (RNN). This innovative amalgamation serves to streamline the training of the GRU and enhances the overall efficiency of the process. In another study, a Bi-directional 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. Shi and Pun [263] combined multiscale local spectral-spatial features obtained through a 3D-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 2D-CNN and Recurrent 3D-CNN for HSIC in [264], proved noteworthy, given the intriguing correlation observed between these methodologies and the respective 2D and 3D-CNN models. This correlation affirms the significance of recurrent CNN in the context of HSIC. In the study [265], CNN and CLSTM were synergistically combined. Firstly, they harnessed the capabilities of a 3D Convolutional Neural Network (CNN) to capture essential low-level spectral-spatial features from the raw hyperspectral data. Subsequently, the researchers employed a Convolutional Long Short-Term Memory (CLSTM) network, a specialized variant of the recurrent neural network, 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 feature extraction process. On the other hand, the second layer is dedicated to acquiring informative features from the HSI, ultimately facilitating accurate classification. Additionally, 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 AEs-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.

Additionally, 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 multi-temporal HSI context, as this could yield significant advancements in various applications. By incorporating recurrent neural networks (RNNs) with HSIC in the analysis of multi-temporal 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. Moreover, as was previously mentioned, the selection of labeled HSI is extremely important and costly due to a number of aspects that either approach for the use of human experts 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

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: i) data wrapping and ii) oversampling [266]. 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 [266], feature space augmentations [267], 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. [268] augmented the training data with three DA operations (flip, rotate, and translation) and then used this augmented data to train CNN for HSIC. In the research study [269], introduced a thorough comparative study of numerous widely used HSI

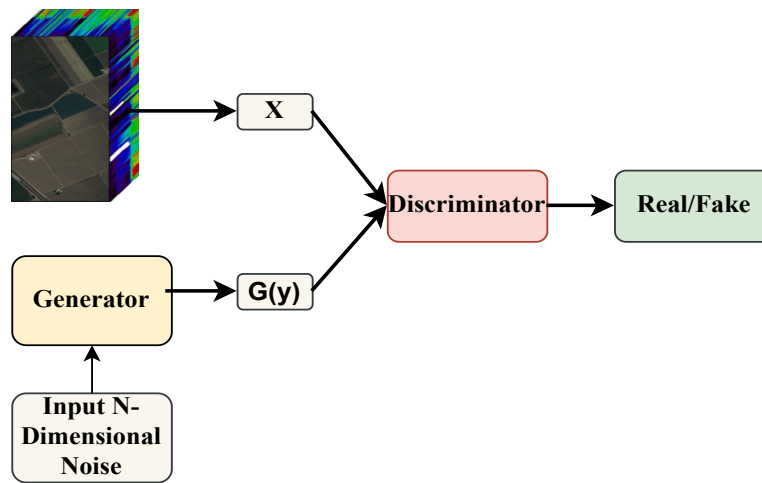


Fig. 11 Illustrates a generic framework of Generative Adversarial Network.

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. In [270], authors assessed the classification efficiency of a Convolutional Neural Network (CNN) in combination with Active Learning (AL) both with and without Data Augmentation (DA) techniques. Their findings revealed that the inclusion of Data Augmentation led to improved classification accuracies, demonstrating its significant contribution to enhancing the model's performance. Consequently, in the study [271], 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 is presented in [272], 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, is utilized for the concluding class label. The study [272] also presented two accelerated DA methods for high-quality data syncretization to enhance the predictive performance of DNN networks. [273], the authors propose a relevant PCA-based online DA approach that reconstructs novel examples during inference rather than training.

B. Semi-Supervised Learning

Semi-supervised 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: Co-training, self-training, GANs, graph-based SSL models, and semi-supervised Support Vector Machines (SVM) are among the commonly employed methods. A Latest thorough review of these SSL methods can be explored in [274]. Furthermore, the authors in [275], also present a comprehensive analysis of SSL methodologies.

There is a brief summary of the SSL-based HSIC methods in [276], where the authors also compare these methodologies in great depth. The approach described in [260] pre-trained a CRNN for HSIC using pseudo or cluster-labeled instances and then fine-tuned the network using limited labeled data. Likewise, Kang et al. [161] presented a semi HSIC approach that uses PCA and increased morphological attribute profiles to retrieve pseudo-labeled sample sizes that are injected into a CNN-based deep feature fusion network.

The research [277] introduced a dual co-training method focused on spectral and spatial features of HSI. Consequently, Zhou et al. [278] pre-trained two SAEs individually, the first pre-trained model utilizing spectral information and the second pre-train model utilizing spatial HSI information, and fine-tuning is obtained through a co-training strategy. To improve the training data, Li et al. [279] introduced a region information-based self-training strategy. In [280], 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. [162] made considerable strides in enhancing the outcomes of HSIC by employing a novel strategy of pseudo-labeling 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 pseudo-labels 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 [281]. 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 few-shot learning 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, the authors in [282] proposed a novel deep few-shot learning method to tackle the limited labeled samples in HSI classification. The method employs three key strategies: using a deep residual 3-D CNN, training the network through episodes to learn a metric space, and using the nearest neighbor classifier within the learned metric space. In another study (), the authors proposed a novel approach to heterogeneous few-shot learning (HFSL) 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 [283]–[286].

Recently, state-of-the-art methods have been proposed with the integration of FSL for HSIC. For example, the authors in [287], propose an innovative classification framework, the Deep Transformer and Few-Shot Learning (DT-FSL), with the aim of achieving fine-grained classification of HSI using only a limited number of instances in a few-shot learning setting. The authors in [288], introduce CTFSL, a novel convolutional transformer-based few-shot learning method. It concurrently applies few-shot learning in source and target domains, establishes a consistent scenario, aligns domain dimensions using a domain aligner, and employs a convolutional transformer (CT) network for local-global feature extraction. In another work in [289], introduce HFC-SST, an advanced spatial-spectral transformer for few-shot hyperspectral image classification. 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. Additionally, a feature extraction network, based on the transformer further exploits spatial-spectral features within the input sequence.

D. Generative Adversarial Network

The GAN presented by Goodfellow et al. [290] consists of two Deep Neural Networks (DNN); the first part of Gan is referred to as a generator, and the second part is called a discriminator, as shown in Figure 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 the paper [291].

Consequently, He et al. [292] presented a spectral-spatial HSIC approach based on GAN. Zhu et al. [293] built CNN-based 1D-GAN and 3D-GAN frameworks to improve classification efficiency for HSIC. The spectral features are generated using a 1D customized GAN [294], which CNN employs for FE, and then a majority voting mechanism is applied to produce HSIC. Feng et al. [295] lately presented a novel Multiclass spatial-spectral GAN (MSGAN), which employs duplet generator models to generate spatial and spectral features with the assistance of several adversarial strategies. Zhong et al. [296] introduced a novel Conditional Random Fields Generative Adversarial Networks (CRFGANs) framework that integrates GAN model with conditional random fields to solve the data sparsity issue for HSIC.

Correspondingly, Wang et al. [297] explored a Caps-Triple-GAN model, which efficiently produces novel sample sizes utilizing a 1D architecture Triple GAN (Triple-GAN) and classifies the produced HSI samples employing Capsule Network (CapsNet). The study [298] introduced using a novel 3D-CNN-based generator network and a 3D deep residual network-based discriminator network for HSIC. In [299], an integration of CapsNet and Convolution LSTM (ConvLSTM)-based discriminator model was presented for HSIC to extract high-level context information.

The study [300] 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. [301] introduced a Generative Adversarial Minority Oversampling-based (GAMO) method to tackle the lasting issue of class-wise data scarcity applied by HSIC to improve predictive accuracy. The authors in [302] propose 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

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. [303] adapted the bottom layers of the pre-trained 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. [304] introduced a proper technique for pretraining and fine-tuning a CNN network before using it to classify novel HSIs. The research work in [305], merged DA and TL techniques to

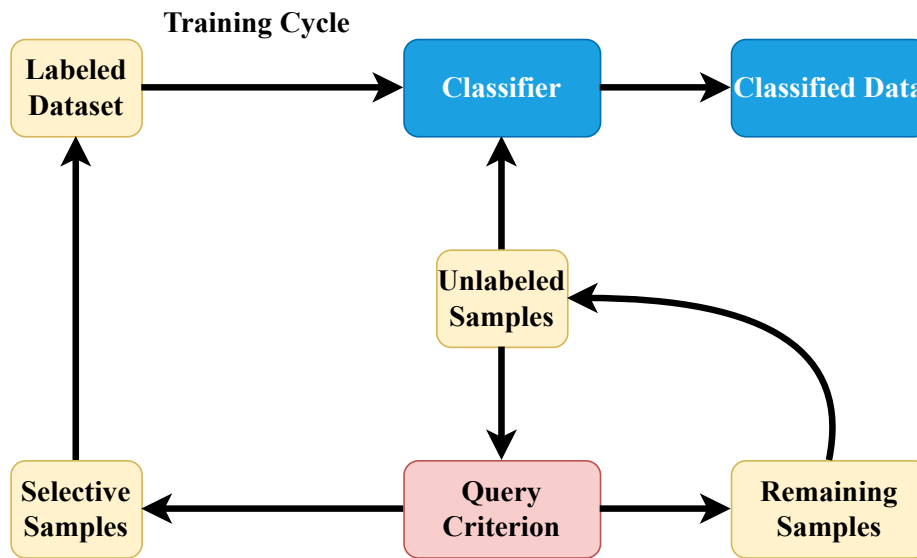


Fig. 12 Shows a generic overview of active learning.

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 [306]). 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. [307] presented an improved heterogeneous TL-based HSIC approach that performs effectively with both homogeneous and heterogeneous HSIs. In a research study [308], the authors introduced a novel heterogeneous TL with a recursive reweighting framework for HSIC. Additionally, the latest article [309] presented a band selection-based TL method to pre-train a CNN that keeps a relatively similar amount of dimensions for different HSIs. Moreover, Lin et al. [310] introduced an unsupervised TL method to classify entirely unexplored target HSI, and de Lima and Marfurt [311] show 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, the authors in [312] introduced SpectralGPT, a pioneering universal RS foundation model specifically designed to process spectral RS images through an innovative 3D generative pretrained transformer (GPT).

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

hyperspectral image classification. 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, the authors in [313], introduce a novel classification framework named Spatial-Spectral Transformer (SST) for hyperspectral image (HSI) 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. The authors in [314], present 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, in [315]–[317], the authors 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, the article [318], introduces 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, in [319], 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. The authors in [320], introduce a novel framework called Multi-Attention Transformer with Adaptive Superpixel Segmentation-Based Active Learning (MAT-ASSAL). This framework demonstrates outstanding classification performance, particularly in scenarios with limited sample sizes. In the study [321], the authors 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 multi-head self-attention (TMHSA).

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, the authors in [322], introduce morphFormer, a novel morphological transformer. The model incorporates a trainable spectral and spatial morphological network, employing morphological convolution operations alongside the attention mechanism. The authors in [323], introduce a Transformer with Residual Upscale GAN (TRUG), consisting of a generator (G) and a discriminator (D). Within the generator (G), a technique called residual upscale (RU) is suggested to enhance the resolution of generated features. Additionally, RU facilitates the extraction of texture features and the capture of contextual relationships. Moreover, the article in [324], introduce TMAC, a Transformer-based Model using Contrastive learning for Enhanced performance (TMAC). The model aims to integrate both methods and enhance overall performance. In [325], The authors 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 (DAU) 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. Figure 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.

The study [326] 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 below.

- 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 [327] or switching the hyperparameters [328]. In general, a smaller number of diversified classification models is sufficient for forming a committee [327], [329].
 - 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 [328].

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 [330].

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 [328].

Recently, AL has significantly empowered Hyperspectral Image Classification (HSIC). Liu et al. [331] introduced the Feature-driven Active Learning (FAL) method, tailored for HSIC, while Zhang et al. [332] proposed Active Semi-supervised Random Forest (ASRF), 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. [333] combined spectral and spatial properties of super-pixels, Xue et al. [334] used neighborhood and super-pixel features to increase data uncertainty, and Bhardwaj et al. [335] 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. [336] 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 [337] introduced a regularized multi-metric batch-mode AL approach for HSIC that takes advantage of several HSI features.

In [338], the authors introduced a Multi-View AL (MVAL) approach that evaluates the instance from diverse perspectives and estimates the information quality of the example using multi-view intensity-based query specifications. In a similar manner, Pradhan et al. [339] used the Fisher discriminant ratio to produce various views, utilizing the approach of multi-learning. In another research study, Zhang et al. [340] introduced a new method of adaptive MVAL approach for HSIC that leverages both spatial and spectral information in each view. Lately, Li et al. [341] presented a new SPSMVAL method that produces different viewpoints for HSIC using

pixel-level, subpixel-level, and super-pixel-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. [342] introduced a novel ADL framework for HSIC, and Sun et al. [343] combined AE with AL methods. Additionally, Haut et al. [344] proposed a novel B-ALCNN method by combining Bayesian CNN with the AL framework for spectral-spatial HSIC. Moreover, Rochac et al. [270] introduced a novel ALCNN approach to effectively leverage unlabeled examples for HSIC. Several studies combined AL and TL for HSIC. In order to accurately reflect the data from the source and target domains, Lin et al. [345] AL-TL approach retrieves the salient examples and makes use of high-level features. In a different study, [346], a stacked sparse AE-based active TL method was put forth that makes use of both spectral and spatial features for HSIC. Another study [347] proposed an Active Multi-kernel Domain Adaptation (AMKDA) 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. [348] introduced MNSAL for multi-class 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. In the work [349], the authors suggest a semi-supervised CNN framework named SA-CNNs, utilizing active learning and superpixel segmentation for hyperspectral image (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 [350], Adaboost [351], and Random Forest [352] 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. Figure 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, [353] presented a multiple classifier framework that integrates an SVM and Random Feature Selection (RFS). In a study [354], 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 [355],

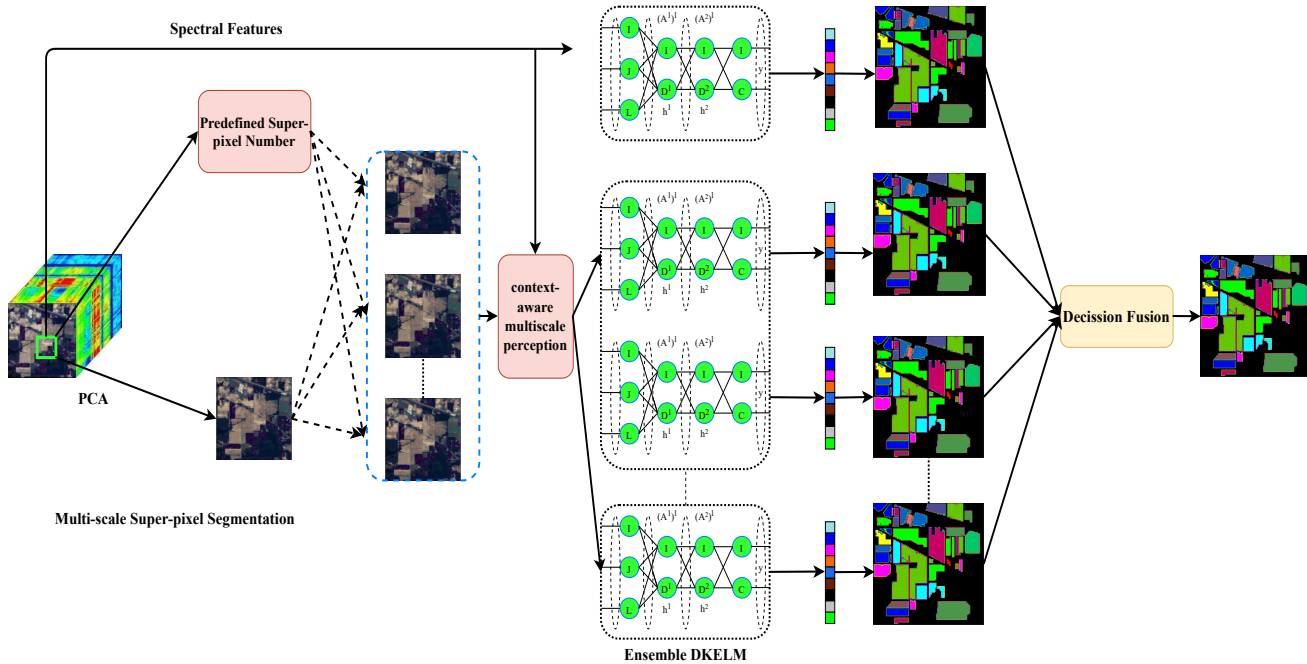


Fig. 13 Illustrates Deep Ensemble Learning Framework.

the authors proposed deep ensemble with the integration of different CNN architectures. Furthermore, they also introduce model augmentation methods that enable the synthesis of novel deep neural networks based on the actual model by integrating the Gaussian noise into the model weights.

In [356], 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 Pixel Pair Feature (PPF) approach to increase the total size of training samples. In [357] proposed a deep ensemble architecture that consists of a deep learning model and sub-space-based ensemble learning. Especially two deep-learning 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 sub-spaces have been employed to diversify the ensemble mechanism in an effective way. Recently, in [358] 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 [359], two methods have been proposed multiview kernel collaborative subspace clustering (MVKCSC) and random subspace MVKCSC (RSMVKCSC). Furthermore, a novel approach to forming the Laplacian matrix employing kernel CR coefficients is presented for clustering based on subspace clustering and CR theory.

In [360], the authors proposed an Enhanced Ensemble Convolution Neural Network (EECNN) method to classify the imbalanced HSI datasets, first they used the random subspace feature method and random forest approach. Then, they built an ensemble learning model by integrating the random feature samples and CNN into the HSI classification. In another

study [361], the authors proposed ensemble CNN with feature subspaces (ECNN-EFSs) method, which used an imbalanced training dataset to train the model and acquired higher accuracy. In [362], proposed spectral-spatial based feature retrieval approach centered on ensemble empirical mode decomposition for HSI classification. In [363], 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.

A study [364] proposed an ensemble margin-based semi-supervised random forest (EMRF) 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. In study [365], proposed a novel ensemble self-supervised feature-learning (ES2FL) technique that instantly extracts deep features conducive to classification without any annotation details, substantially lowering deep-learning 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 ap-

proaches, including feature level and view level ensemble. In article [366], proposed an Ensemble Stochastic Watershed Edge Weights (ESWEW) method. The introduced method consists of two main points. First, the ensemble approach reduces the variance, and second categories in HSI datasets and feature similarity have only one-sided implications. In article [367], introduced a novel evolutionary multitask ensemble learning model (EMT EL). Firstly, 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. Secondly, 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 feature extraction, 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 in [368], [369] used DF for HSIC to acquire enhanced performance. Advanced DF methods have been proposed to tackle the HSIC efficiency problem [370], [371]. These methods integrate the cascade method in DF to boost the performance without facing the computational complexity issue. The authors in [372], introduce a novel deep model, named Densely Connected Deep Random Forest (DCDRF), designed for hyperspectral image classification. The model comprises multiple forward-connected random forests. Recently, in [373], the authors 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-multi-grained scanning features are fed into a deep forest classifier to accomplish the classification task.

Another study [374], 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, ensembles the training models with their weights for the classification purpose. In a study [375], proposed a novel flexible ensemble method for the multiclass imbalance issue focused on SMOTE and RoF with diversified sampling rates (AdaSRoF). The presented method uses SMOTE and a dynamic information sampling ratio for core classifiers to generate a series of balanced data sets with further diversification and much less noise. In [376], proposed a label propagation ensemble (LPE) method.

In LPE, a random subspace 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 pseudo labels. During the iterative process, extreme learning machine models are given training on labeled and pseudo-labeled instances. In [377], introduce a Multiscale Convolutional Neural Networks (CNNs) Ensemble Based Self-Learning (MCE-SL) technique for semi-supervised 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 data set for the next training iteration.

In [378], proposed multi-view-based random rotation ensemble pruning (MVRR-EP) and presented some state-of-the-art features. Firstly, 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. Secondly, 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. Thirdly, 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. Another article [379] presented the random subspace ensemble with improved feature (RSE-EF), 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 (SNSS) technique. Following that, they generate improved random subspaces (ERSs) with reduced dimensionality and more unique features than the actual random subspaces (RSs) in order to mitigate the curse of a high feature-to-instance ratio more efficiently.

In [380], the authors introduced an adaptive semi-supervised rotation forest (SSRoF) for hyperspectral image classification with sparse training data. The proposal relies on Rotation Forest (RoF), a classification method that has proven to be useful in the case of high-dimensional data. It is altered for semi-supervised learning 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 (OOB) examples are employed to determine the efficient number of instances to include in the training set. In [381], 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 [382], the vicinity of class noise, a new two-phase ensemble-based data filtering technique is presented to enhance hyperspectral image classification performance. The presented technique

combines noise redundancy classification models with sensitive techniques.

Recently, in article [383] introduced two novel dynamic ensemble learning techniques employing local weighted residual (LWR-DEL) and double-weighted residual (DWR-DEL) of multi-collaborative representation classifiers. Initially, the dynamic ensemble learning technique focused on clustering is employed to present initial information for the collaborative representation 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 K-nearest neighbor 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. In the article [384], the authors introduce a novel multimodal deep learning framework for land use and land cover (LULC) classification. They achieve this by making minimal modifications to the conventional Vision Transformer (ViT). In the study [385], the authors investigate the potential of unsupervised cross-modal feature representation learning. They achieve this by incorporating multimodal data into a fully recombined matrix format. The article [386], presents two innovative network components: a self-generative adversarial networks (GANs) module and a mutual-GANs module. These additions aim to learn perturbation-insensitive feature representations and bridge the gap between modalities, enhancing effective and robust feature transfer. The article [387], proposes enhancing feature extraction 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. Although, it can significantly maximize the input sequence for RNNs, which can lead to an overfitting problem.

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. Additionally, 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 semi-supervised learning and transfer learning 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. Although 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 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. Although, 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 [388]. The Salinas scene is made up of 224 spectral bands composed of the wavelength range of 360 to 2500nm and 512×217 pixels, as shown in Table I. The data set 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 data set.

TABLE I
DETAILED DESCRIPTION OF EACH DATASET USED DURING EXPERIMENT

Name	Spatial Dimension	Spectral Bands	Wavelength Range	Classes
IP	145x145	224	400nm - 2500nm	16
UP	610x340	103	430nm - 860nm	9
SA	512x217	224	360nm - 2500nm	16
KSC	610x340	224	400nm - 2500nm	13

The second dataset PU was collected by an unmanned aerial optical device, the Reflective Optics Spectrographic Imaging System over Pavia, Italy, 2002 (ROSIS) [389]. 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 data set mainly comprises 640×340 pixels with a 1.3 m/pixel spatial resolution and 103 bands containing the wavelength range of 430 to 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 third dataset is the Purdue Indiana Indian Pines scene IP [390], 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 Table I, and IV.

The fourth Kennedy Space Center (KSC) dataset was collected in 1996 by AVIRIS [391], with wavelengths ranging from 400-2500 nm. The image of the KSC dataset contains 512×614 pixels and 176 spectral bands after the elimination 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. Figure 16 demonstrates the comprehensive class detail and ground truth maps. Furthermore, the disjoint training/test sample maps are shown in Figure 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, remain constant across all experimental techniques. This uniformity ensures a level playing field and a fair comparison of results, thereby enhancing the reliability and validity of our findings.

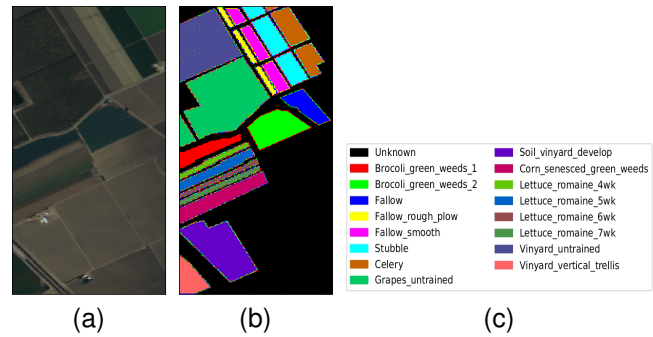


Fig. 14 Shows the SA dataset. (a) source image, and (b) ground truth.

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

Class No.	Class Name	Training	Test
1	Broccoli_green_weeds_1	201	1808
2	Broccoli_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

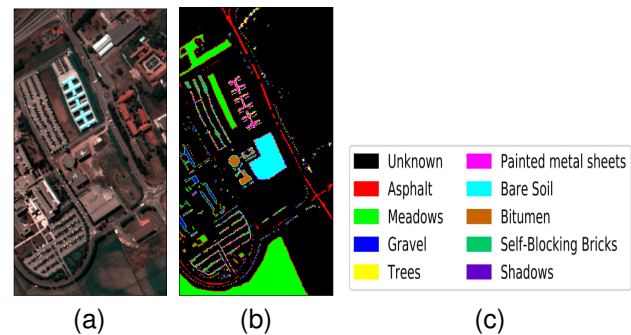


Fig. 15 Shows the PU dataset. (a) source image, and (b) ground truth.

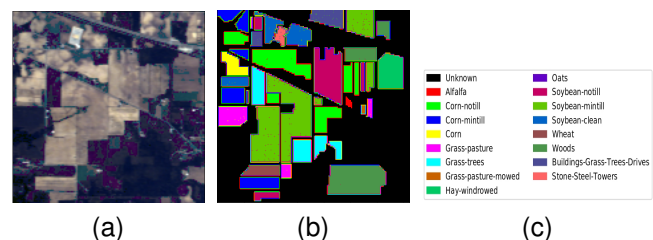


Fig. 16 Shows the IP dataset. (a) source image, and (b) ground truth.

TABLE III
AMOUNTS 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

TABLE IV
AMOUNTS 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

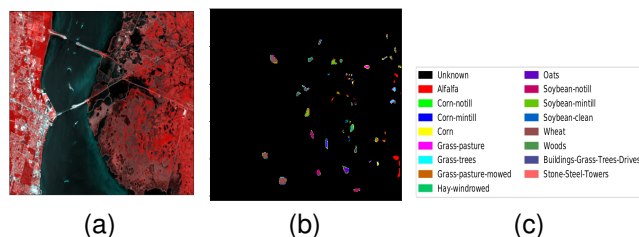


Fig. 17 Shows the KSC dataset. (a) source image, and (b) ground truth.

TABLE V
AMOUNTS OF TRAINING AND TEST DATA OF KSC SCENE DATASET

Class No.	Class Name	Training	Test
1	CP hammock	5	41
2	Hardwood	143	1285
3	Spartina marsh	83	747
4	Mud flats	24	213
5	Scrub	49	434
6	Slash pine	73	657
7	Swap	3	25
8	Cattail marsh	48	430
9	Water	2	18
10	Willom swamp	98	874
11	Oak/broadleaf	246	2209
12	Graminoid marsh	60	533
13	Salt marsh	21	184
Total		1031	9218

B. Experimental Results and Discussion

The most significant advancements in recent years have been produced in the areas of HSIC, SVM-RBF¹, CFF-200², 2D-CNN [392], GCNN [393], FADCNN [394], NL-GCNN [395], DECNN [396], MCE-SL [397], TCRC-Boosting [398], EMT-EL [399], ECNN-PPF-SSF [400], and DHS [401] 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 deep learning based approaches for HSIC (DECNN), multi-class 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.

All of the aforementioned research study to certain 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 of 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, VII, VIII, and IX and Figures 18, 19, 20, and 21 demonstrate the acquired accuracies for disjoint training and test samples. The 10-cross-validation operation was used to analyze the Overall Accuracy (OA%), Average Accuracy (AA%), and Kappa (k) accuracy for comparative processes in all of the results depicted in the Tables and Figures. For this case specifically, let us suppose the

¹<https://www.csie.ntu.edu.tw/~cjlin/libsvm/>

²<https://github.com/twgr/ccfs>

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

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
K	-	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, KAPPA ON THE SALINAS DATA SET

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, KAPPA ON THE INDIAN PINES DATA SET

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
K	70.93	80.59	89.18	83.42	86.10	86.25	88.50	89.23	90.06	91.32	91.96	96.26

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

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

case of Salinas scene results; the work [401] has the highest OA, AA, and k accuracies which are 99.96%, 99.97%, and 99.95%, respectively, in comparison with the OA, AA, and k accuracies for ensemble-based methods comparative results; 99.18%, 99.58%, and 98.17% for [400], 99.17%, 99.45%, 98.11% for [399], 99.10%, 99.45%, 97.56% for [398], 99.08%, 99.54%, 98.23% for [397], 98.84%, 99.37%, 97.15% for [396], and for non-ensemble methods are 92.28%, 96.39%, 89.64% for [395], 90.58%, 94.56%, 89.28% for [394], 90.25%, 93.69%, 89.18% for [393], 89.72%, 95.43%, 88.58% for CFF-200, and 88.82%, 94.67%, 87.57% for SVM-RBF. Similar findings were produced with the other experimental datasets. Tables VI, VII, VIII, and IX highlight the increased accuracies in bold.

As illustrated in Table II and Figure 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.

In overall, the study [401], [400], and [399] 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. Additionally, the accuracy of such techniques improves with the number of training samples. Although, 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, since analyzing disjoint train/test samples, the studies [401], [400], and [399] 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 the samples to be labeled; when there are no limitations, such techniques may

learn nothing. Additionally, CNN has a symmetric architecture that also results in excess training parameters, increasing the complexity of training. The studies [395], and [394] mitigate the aforementioned problems, although the research study [393] does not use the greedy layer-wise 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 performance 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, Graph-based CNN impacts the choice of the most essential samples for training, allowing the model to emphasize more focus on identical samples for HSIC, although Few-Shot Learning (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 the 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 CNN

This section covered 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 feature extraction methods such as Makantasis et al. [392] 2D convolution neural network for HSIC (2D-CNN) and D. hong et al. [393] Graph Convolutional Neural Network (GCNN) for hyperspectral image classification (GCNN) and Lately, Roy et al. [394] Feedback Attention-based Dense Convolutional Neural Network (FADCNN) 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.

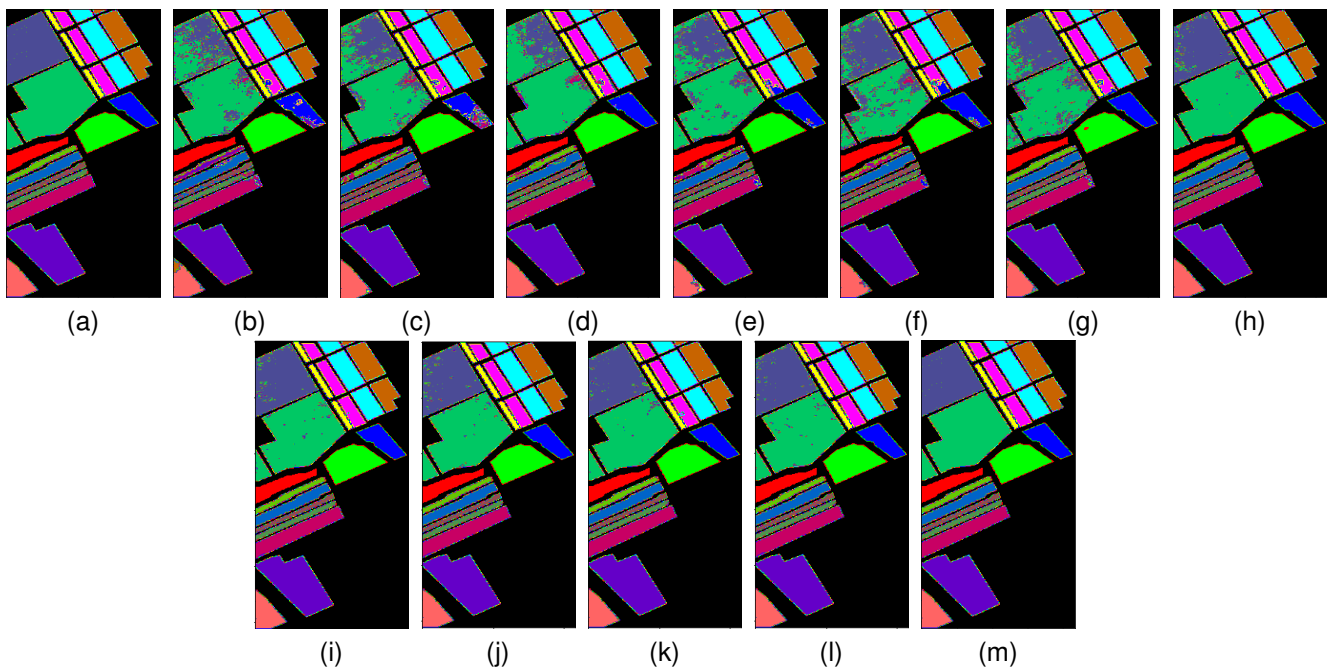


Fig. 18 Shows the classification maps of various methods for the Salinas data set. (From Left to Right) (a) ground truth, (b) SVM-RBF, (c) CCF-200, (d) 2-D CNN, (e) GCNN, (f) FADCNN, (g) NL-GCNN, (h) DECNN, (i) MCE-SL, (j) TCRC-Boosting, (k) EMT-EL, (l) ECNN-PPF-SSF, and (m) DHS.

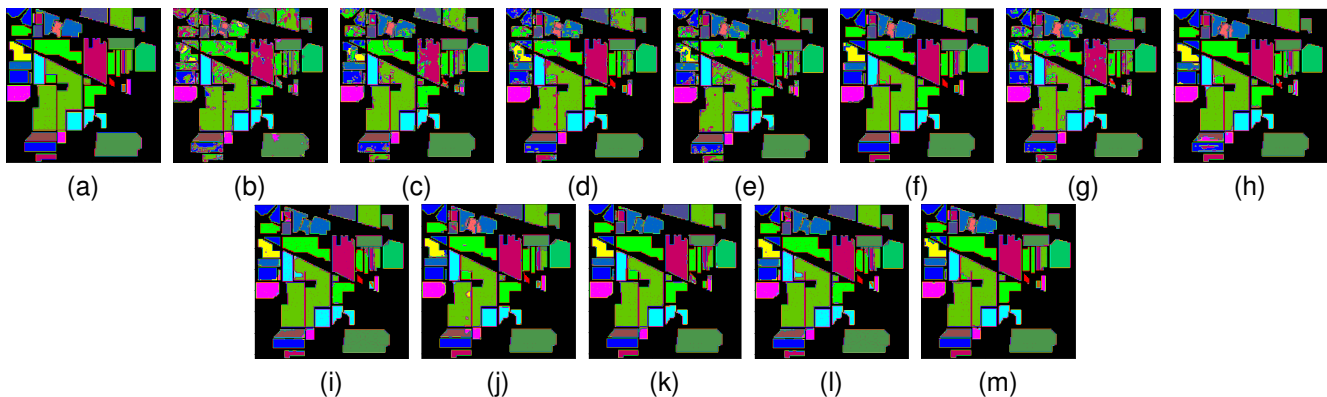


Fig. 19 Shows the classification maps of various methods for the Indian Pines data set. (From Left to Right and from top to bottom) (a) ground truth, (b) SVM-RBF, (c) CCF-200, (d) 2-D CNN, (e) GCNN, (f) FADCNN, (g) NL-GCNN, (h) DECNN, (i) MCE-SL, (j) TCRC-Boosting, (k) EMT-EL, (l) ECNN-PPF-SSF and (m)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

All of the previously mentioned CNN feature extraction methods have been tested on four various hyperspectral datasets, such as the IP, PU, SA, and KSC datasets. Table VI, VII, VIII, and IX demonstrates the experimental accuracy. From all of these analyses, it is evident that the FADCNN method showed better performance than 2D-CNN, GCNN, and NL-GCNN. All datasets have shown a typical pattern, where the GCNN and NL-GCNN methods outperformed the conventional 2D-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 2D-CNN to the 3D-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 paper, we surveyed novel approaches in HSIC utilizing cutting-edge DNNs [for example, AE, DBN, RNN, CNN, TL, FSL, active/self-learning (AL/SL), and DA], ensemble and deep ensemble in a wide range of learning approaches. Moreover, we explored techniques for overcoming the difficulties of the

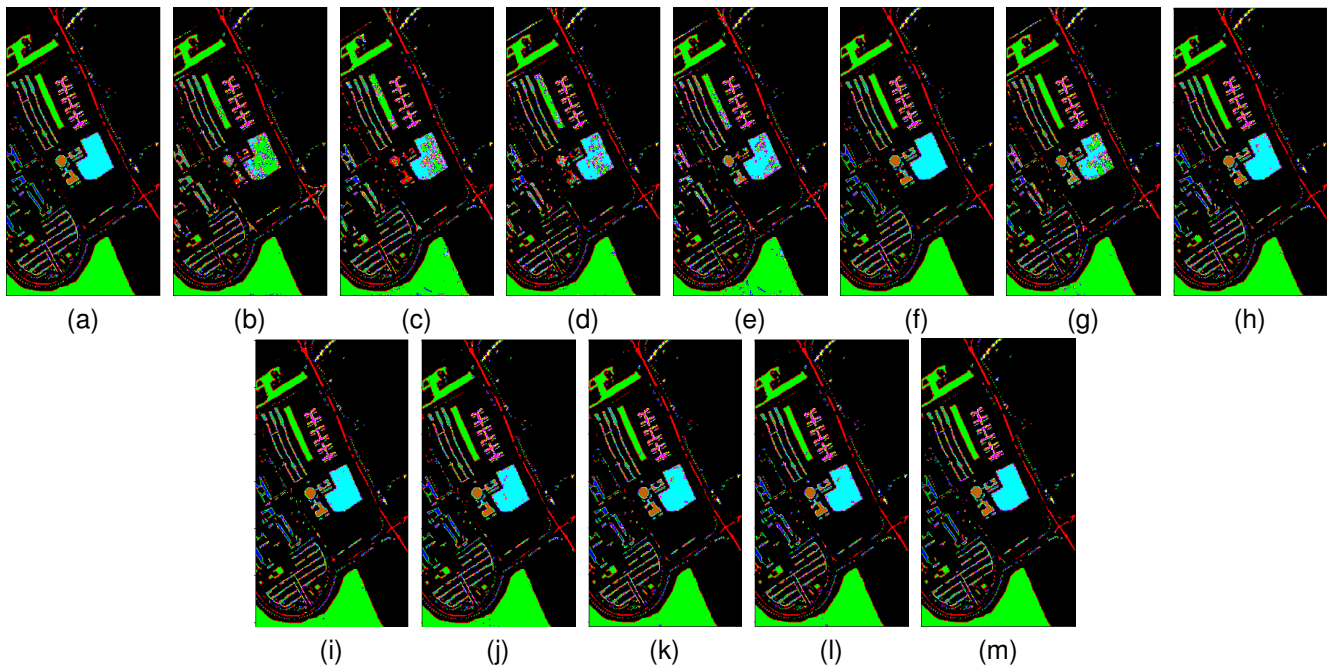


Fig. 20 Shows the classification maps of various methods for the Pavia University data set. (From Left to Right) (a) ground truth, (b) SVM-RBF, (c) CCF-200, (d) 2-D CNN, (e) GCNN, (f) FADCNN, (g) NL-GCNN, (h) DECNN, (i) MCE-SL, (j) TCRC-Boosting, (k) EMT-EL, (l) ECNN-PPF-SSF, and (m) DHS.

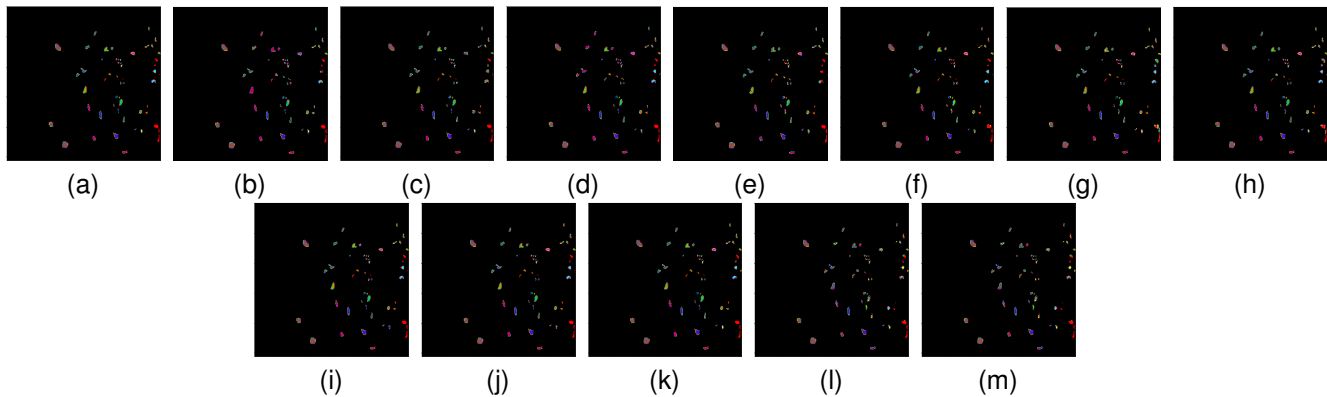


Fig. 21 Shows the classification maps of various methods for the KSC data set. (From Left to Right) (a) ground truth, (b) SVM-RBF, (c) CCF-200, (d) 2-D CNN, (e) GCNN, (f) FADCNN, (g) NL-GCNN, (h) DECNN, (i) MCE-SL, (j) TCRC-Boosting, (k) EMT-EL, (l) ECNN-PPF-SSF, and (m) DHS.

sparsity of training data accessibility, such as DA, FSL, TL, and AL. We chose several comparative studies to perform the experimentations on benchmark HSI datasets using the techniques mentioned in the previous section.

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 is 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 deep learning community. They can be considered part of the data augmentation 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 data augmentation 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 becomes available. Incremental learning holds significant potential for HSIC, providing a framework for continuous adaptation and improvement as new hyperspectral data becomes available. In the context of hyperspectral imaging, 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, D. Toomey, P. Onaka, A. Denault, W. Stahlberger, W. Vacca, M. Cushing, and S. Wang, "Spex: a medium-resolution 0.8–5.5 micron spectrograph and imager for the nasa infrared telescope facility," *Publications of the Astronomical Society of the Pacific*, vol. 115, no. 805, p. 362, 2003.
- [2] M. Ahmad, A. Khan, A. M. Khan, M. Mazzara, S. Distefano, A. Sohaib, and O. Nibouche, "Spatial prior fuzziness pool-based interactive classification of hyperspectral images," *Remote Sensing*, vol. 11, no. 9, p. 1136, 2019.
- [3] D. Hong, W. He, N. Yokoya, J. Yao, L. Gao, L. Zhang, J. Chanussot, and X. Zhu, "Interpretable hyperspectral artificial intelligence: When nonconvex modeling meets hyperspectral remote sensing," *IEEE Geoscience and Remote Sensing Magazine*, vol. 9, no. 2, pp. 52–87, 2021.
- [4] H. Ayaz, M. Ahmad, A. Sohaib, M. N. Yasir, M. A. Zaidan, M. Ali, M. H. Khan, and Z. Saleem, "Myoglobin-based classification of minced meat using hyperspectral imaging," *Applied Sciences*, vol. 10, no. 19, p. 6862, 2020.
- [5] M. H. Khan, Z. Saleem, M. Ahmad, A. Sohaib, H. Ayaz, and M. Mazzara, "Hyperspectral imaging for color adulteration detection in red chili," *Applied Sciences*, vol. 10, no. 17, p. 5955, 2020.
- [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. 176 986–176 996, 2020.
- [7] M. Zulfiqar, M. Ahmad, A. Sohaib, M. Mazzara, and S. Distefano, "Hyperspectral imaging for bloodstain identification," *Sensors*, vol. 21, no. 9, p. 3045, 2021.
- [8] H. Ayaz, M. Ahmad, M. Mazzara, and A. Sohaib, "Hyperspectral imaging for minced meat classification using nonlinear deep features," *Applied Sciences*, vol. 10, no. 21, p. 7783, 2020.
- [9] M. H. Khan, Z. Saleem, M. Ahmad, A. Sohaib, H. Ayaz, M. Mazzara, and R. A. Raza, "Hyperspectral imaging-based unsupervised adulterated red chili content transformation for classification: Identification of red chili adulterants," *Neural Computing and Applications*, vol. 33, no. 21, pp. 14 507–14 521, 2021.
- [10] "10 important applications of hyperspectral image," <https://grindgis.com/remote-sensing/10-important-applications-of-hyperspectral-image>, accessed: 2023-05-30.
- [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," *Critical reviews in food science and nutrition*, vol. 59, no. 1, pp. 173–180, 2019.
- [12] M. Ahmad, "Ground truth labeling and samples selection for hyperspectral image classification," *Optik*, vol. 230, p. 166267, 2021.
- [13] "hyperspectral image classification application," <https://resonon.com/applications>, accessed: 2023-05-30.
- [14] M. Ahmad, A. M. Khan, and R. Hussain, "Graph-based spatial-spectral feature learning for hyperspectral image classification," *IET image processing*, 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," *Journal of Real-Time Image Processing*, 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 journal of photogrammetry and remote sensing*, 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," *International Journal of Soft Computing and Engineering (IJSCE)*, 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 Transactions on Cybernetics*, vol. 51, no. 7, pp. 3602–3615, 2020.
- [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 transactions on geoscience and remote sensing*, vol. 49, no. 7, pp. 2635–2646, 2011.
- [20] J. M. Bioucas-Dias, A. Plaza, N. Dobigeon, M. Parente, Q. Du, P. Gader, and J. Chanussot, "Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches," *IEEE journal of selected topics in applied earth observations and remote sensing*, vol. 5, no. 2, pp. 354–379, 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 Journal of Photogrammetry and Remote Sensing*, vol. 119, pp. 49–63, 2016.
- [23] M. Ahmad and I. U. Haq, "Linear unmixing and target detection of hyperspectral imagery using osp," *proc. of 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 *2011 International Conference on Information Communication and Management, IACSIT Press*, vol. 10, 2011, 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 Transactions on Image Processing*, vol. 28, no. 4, pp. 1923–1938, 2018.
- [27] L. Gao, Z. Han, D. Hong, B. Zhang, and J. Chanussot, “Cycu-net: Cycle-consistency unmixing network by learning cascaded autoencoders,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–14, 2021.
- [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 processing magazine*, vol. 19, no. 1, pp. 58–69, 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 Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 10, no. 9, pp. 4124–4137, 2017.
- [30] S. Li, K. Zhang, Q. Hao, P. Duan, and X. Kang, “Hyperspectral anomaly detection with multiscale attribute and edge-preserving filters,” *IEEE Geoscience and Remote Sensing Letters*, vol. 15, no. 10, pp. 1605–1609, 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 Transactions on Geoscience and Remote Sensing*, vol. 57, no. 7, pp. 5146–5158, 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 Geoscience and Remote Sensing Magazine*, vol. 7, no. 2, pp. 140–158, 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 Geoscience and Remote Sensing Letters*, vol. 17, no. 2, pp. 302–306, 2019.
- [34] P. Chen, B. Zhang, D. Hong, Z. Chen, X. Yang, and B. Li, “Fccdn: Feature constraint network for vhr image change detection,” *ISPRS Journal of Photogrammetry and Remote Sensing*, 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,” *Proceedings of the IEEE*, vol. 101, no. 3, pp. 652–675, 2012.
- [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, p. e0188996, 2018.
- [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, p. 163712, 2020.
- [38] M. Ahmad, A. M. Khan, R. Hussain, S. Protasov, F. Chow, and A. M. Khattak, “Unsupervised geometrical feature learning from hyperspectral data,” in *2016 IEEE Symposium Series on Computational Intelligence (SSCI)*. IEEE, 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,” *CoRR*, vol. abs/1705.06920, 2017.
- [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 Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 464–473, 2020.
- [42] D. Hong, N. Yokoya, J. Chanussot, and X. X. Zhu, “Cospace: Common subspace learning from hyperspectral-multispectral correspondences,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 7, pp. 4349–4359, 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 Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 6, no. 2, pp. 488–498, 2013.
- [44] W. Wei, L. Zhang, C. Tian, A. Plaza, and Y. Zhang, “Structured sparse coding-based hyperspectral imagery denoising with intracluster filtering,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 12, pp. 6860–6876, 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 Transactions on Geoscience and Remote Sensing*, vol. 55, no. 7, pp. 3823–3834, 2017.
- [46] C. Yi, Y.-Q. Zhao, and J. C.-W. Chan, “Hyperspectral image super-resolution based on spatial and spectral correlation fusion,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 7, pp. 4165–4177, 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 Transactions on Geoscience and Remote Sensing*, vol. 53, no. 8, pp. 4238–4249, 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 Geoscience and Remote Sensing Letters*, vol. 13, no. 6, pp. 747–751, 2016.
- [49] H. Wu, B. Liu, W. Su, W. Zhang, and J. Sun, “Hierarchical coding vectors for scene level land-use classification,” *Remote Sensing*, vol. 8, no. 5, p. 436, 2016.
- [50] G. Cheng, J. Han, and X. Lu, “Remote sensing image scene classification: Benchmark and state of the art,” *Proceedings of the IEEE*, vol. 105, no. 10, pp. 1865–1883, 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 journal of photogrammetry and remote sensing*, vol. 147, pp. 193–205, 2019.
- [52] T. R. Martha, N. Kerle, C. J. Van Westen, V. Jetten, and K. V. Kumar, “Segmentation and data-driven thresholding for knowledge-based landslide detection by object-based image analysis,” *IEEE transactions on geoscience and remote sensing*, vol. 49, no. 12, pp. 4928–4943, 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 bow and pls,” *International Journal of Remote Sensing*, 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,” *International Journal of Remote Sensing*, 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,” *International journal of remote sensing*, 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,” *Artificial Intelligence Review*, vol. 26, no. 3, pp. 159–190, 2006.
- [57] S. B. Kotsiantis, I. Zaharakis, P. Pintelas *et al.*, “Supervised machine learning: A review of classification techniques,” *Emerging artificial intelligence applications in computer engineering*, 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 Transactions on Geoscience and Remote Sensing*, vol. 59, no. 10, pp. 8568–8583, 2020.
- [60] A. Plaza, J. A. Benediktsson, J. W. Boardman, J. Brazile, L. Bruzzone, G. Camps-Valls, J. Chanussot, M. Fauvel, P. Gamba, A. Gualtieri *et al.*, “Recent advances in techniques for hyperspectral image processing,” *Remote sensing of environment*, vol. 113, pp. S110–S122, 2009.
- [61] R. Ablin and C. H. Sulochana, “A survey of hyperspectral image classification in remote sensing,” *International Journal of Advanced Research in Computer and Communication Engineering*, 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 processing magazine*, vol. 31, no. 1, pp. 45–54, 2013.
- [63] D. Chutia, D. Bhattacharyya, K. K. Sarma, R. Kalita, and S. Sudhakar, “Hyperspectral remote sensing classifications: a perspective survey,” *Transactions in 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 Journal of Selected topics in applied earth observations and remote sensing*, vol. 7, no. 6, pp. 2094–2107, 2014.
- [65] P. Ghamisi, N. Yokoya, J. Li, W. Liao, S. Liu, J. Plaza, B. Rasti, and A. Plaza, “Advances in hyperspectral image and signal processing: A comprehensive overview of the state of the art,” *IEEE Geoscience and Remote Sensing Magazine*, vol. 5, no. 4, pp. 37–78, 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, p. 204, 2019.
- [67] B. Rasti, D. Hong, R. Hang, P. Ghamisi, X. Kang, J. Chanussot, and J. A. Benediktsson, "Feature extraction for hyperspectral imagery: The evolution from shallow to deep: Overview and toolbox," *IEEE Geoscience and Remote Sensing Magazine*, vol. 8, no. 4, pp. 60–88, 2020.
- [68] H. Petersson, D. Gustafsson, and D. Bergstrom, "Hyperspectral image analysis using deep learning—a review," in *2016 sixth international conference on image processing theory, tools and applications (IPTA)*. IEEE, 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 Geoscience and Remote Sensing Letters*, vol. 19, pp. 1–5, 2020.
- [70] D. Hong, Z. Han, J. Yao, L. Gao, B. Zhang, A. Plaza, and J. Chanussot, "Spectralformer: Rethinking hyperspectral image classification with transformers," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–15, 2021.
- [71] M. Paoletti, J. Haut, J. Plaza, and A. Plaza, "Deep learning classifiers for hyperspectral imaging: A review," *ISPRS Journal of Photogrammetry and Remote Sensing*, 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 Transactions on Geoscience and Remote Sensing*, vol. 57, no. 8, pp. 5384–5394, 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 Sensing*, vol. 8, no. 6, p. 483, 2016.
- [74] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*, vol. 1. Ieee, 2005, pp. 886–893.
- [75] A. Oliva and A. Torralba, "Modeling the shape of the scene: A holistic representation of the spatial envelope," *International journal of computer vision*, vol. 42, pp. 145–175, 2001.
- [76] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proceedings of the seventh IEEE international conference on computer vision*, vol. 2. Ieee, 1999, 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 Transactions on Geoscience and Remote Sensing*, vol. 43, no. 3, pp. 492–501, 2005.
- [78] G. Camps-Valls and L. Bruzzone, "Kernel-based methods for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 6, pp. 1351–1362, 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 journal on advances in signal processing*, 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 *2016 4th International Workshop on Earth Observation and Remote Sensing Applications (EORSA)*. IEEE, 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 Computer Vision*, vol. 9, no. 5, pp. 639–647, 2015.
- [82] R. Azhar, D. Tuwohingide, D. Kamudi, N. Suciati *et al.*, "Batik image classification using sift feature extraction, bag of features and support vector machine," *Procedia Computer Science*, vol. 72, pp. 24–30, 2015.
- [83] O. Zeglazi, A. Amine, and M. Rziza, "Sift descriptors modeling and application in texture image classification," in *2016 13th International Conference on Computer Graphics, Imaging and Visualization (CGiV)*. IEEE, 2016, pp. 265–268.
- [84] Y. Xu, K. Hu, Y. Tian, and F. Peng, "Classification of hyperspectral imagery using sift for spectral matching," in *2008 Congress on Image and Signal Processing*, vol. 2. IEEE, 2008, pp. 704–708.
- [85] Y. Yang and S. Newsam, "Comparing sift descriptors and gabor texture features for classification of remote sensed imagery," in *2008 15th IEEE international conference on image processing*. IEEE, 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 *2019 26th international conference on telecommunications (ICT)*. IEEE, 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," *Digital Signal Processing*, vol. 82, pp. 152–165, 2018.
- [88] R. M. Haralick, K. Shanmugam, and I. H. Dinstein, "Textural features for image classification," *IEEE Transactions on systems, man, and cybernetics*, no. 6, pp. 610–621, 1973.
- [89] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 24, no. 7, pp. 971–987, 2002.
- [90] L. Zhao, P. Tang, and L. Huo, "Feature significance-based multibag-of-visual-words model for remote sensing image scene classification," *Journal of Applied Remote Sensing*, vol. 10, no. 3, pp. 035 004–035 004, 2016.
- [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 Geoscience and Remote Sensing Letters*, vol. 10, no. 5, pp. 1055–1059, 2013.
- [92] Y. Yang and S. Newsam, "Bag-of-visual-words and spatial extensions for land-use classification," in *Proceedings of the 18th SIGSPATIAL international conference on advances in geographic information systems*, 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 Geoscience and Remote Sensing Letters*, vol. 7, no. 2, pp. 366–370, 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 Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 9, no. 6, pp. 2343–2353, 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 Geoscience and Remote Sensing Letters*, vol. 12, no. 6, pp. 1357–1361, 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 Journal of Photogrammetry and Remote Sensing*, 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 Sensing*, vol. 8, no. 2, p. 157, 2016.
- [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 Sensing*, vol. 7, no. 11, pp. 14 988–15 013, 2015.
- [99] S. Lazebnik, C. Schmid, and J. Ponce, "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories," in *2006 IEEE computer society conference on computer vision and pattern recognition (CVPR'06)*, vol. 2. IEEE, 2006, 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 Transactions on Geoscience and Remote Sensing*, vol. 54, no. 4, pp. 2108–2123, 2015.
- [101] R. Kusumaningrum, H. Wei, R. Manurung, and A. Murni, "Integrated visual vocabulary in latent dirichlet allocation-based scene classification for ikonos image," *Journal of Applied Remote Sensing*, vol. 8, no. 1, pp. 083 690–083 690, 2014.
- [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 Transactions on Geoscience and Remote Sensing*, vol. 53, no. 11, pp. 6207–6222, 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 Journal of Photogrammetry and Remote Sensing*, 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 Sensing*, vol. 8, no. 3, p. 259, 2016.
- [105] V. Risojević and Z. Babić, "Fusion of global and local descriptors for remote sensing image classification," *IEEE Geoscience and Remote Sensing Letters*, vol. 10, no. 4, pp. 836–840, 2012.
- [106] M. L. Mekhalfi, F. Melgani, Y. Bazi, and N. Alajlan, "Land-use classification with compressive sensing multifeature fusion," *IEEE Geoscience and Remote Sensing Letters*, vol. 12, no. 10, pp. 2155–2159, 2015.
- [107] G. Sheng, W. Yang, T. Xu, and H. Sun, "High-resolution satellite scene classification using a sparse coding based multiple feature combination," *International journal of remote sensing*, 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, L. Gao, N. Yokoya, J. Yao, J. Chanussot, Q. Du, and B. Zhang, "More diverse means better: Multimodal deep learning meets remote-sensing imagery classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 5, pp. 4340–4354, 2020.
- [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 Geoscience and Remote Sensing Letters*, vol. 12, no. 11, pp. 2321–2325, 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 Sensing*, vol. 7, no. 11, pp. 14 680–14 707, 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 Transactions on Geoscience and Remote Sensing*, 2023.
- [114] S. Chen and Y. Wang, "Convolutional neural network and convex optimization," *Dept. of Elect. and Comput. Eng., Univ. of California at San Diego, San Diego, CA, USA, Tech. Rep.*, 2014.
- [115] P. Hammer, "Adaptive control processes: a guided tour (r. bellman)," 1962.
- [116] G. Hughes, "On the mean accuracy of statistical pattern recognizers," *IEEE transactions on information theory*, vol. 14, no. 1, pp. 55–63, 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 Geoscience and remote sensing magazine*, vol. 1, no. 2, pp. 6–36, 2013.
- [119] Q. Nguyen and M. Hein, "Optimization landscape and expressivity of deep cnns," in *International conference on machine learning*. PMLR, 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, p. 167757, 2021.
- [121] M. Ahmad, M. Mazzara, and S. Distefano, "Regularized cnn feature hierarchy for hyperspectral image classification," *Remote Sensing*, vol. 13, no. 12, p. 2275, 2021.
- [122] L. Bottou *et al.*, "Stochastic gradient learning in neural networks," *Proceedings of Neuro-Nimes*, vol. 91, no. 8, p. 12, 1991.
- [123] N. Qian, "On the momentum term in gradient descent learning algorithms," *Neural networks*, 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," *arXiv preprint arXiv:1412.6980*, 2014.
- [126] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," *arXiv preprint arXiv:1711.05101*, 2017.
- [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 transactions on neural networks and learning systems*, vol. 31, no. 11, pp. 4500–4511, 2019.
- [128] L. Liu, H. Jiang, P. He, W. Chen, X. Liu, J. Gao, and J. Han, "On the variance of the adaptive learning rate and beyond," *arXiv preprint arXiv:1908.03265*, 2019.
- [129] H. Yong, J. Huang, X. Hua, and L. Zhang, "Gradient centralization: A new optimization technique for deep neural networks," in *Computer Vision—ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part I 16*. Springer, 2020, pp. 635–652.
- [130] S. K. Roy, M. E. Paoletti, J. M. Haut, S. R. Dubey, P. Kar, A. Plaza, and B. B. Chaudhuri, "Angulargrad: A new optimization technique for angular convergence of convolutional neural networks," *arXiv preprint arXiv:2105.10190*, 2021.
- [131] D. Erhan, A. Courville, Y. Bengio, and P. Vincent, "Why does unsupervised pre-training help deep learning?" in *Proceedings of the thirteenth international conference on artificial intelligence and statistics*. JMLR Workshop and Conference Proceedings, 2010, pp. 201–208.
- [132] M. Z. Alom, T. M. Taha, C. Yakopcic, S. Westberg, P. Sidike, M. S. Nasrin, M. Hasan, B. C. Van Essen, A. A. Awwal, and V. K. Asari, "A state-of-the-art survey on deep learning theory and architectures," *electronics*, vol. 8, no. 3, p. 292, 2019.
- [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 Computing*, 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 Processing Magazine*, vol. 28, no. 3, pp. 119–126, 2011.
- [135] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
- [136] Y. Bengio, P. Simard, and P. Frasconi, "Learning long-term dependencies with gradient descent is difficult," *IEEE transactions on neural networks*, vol. 5, no. 2, pp. 157–166, 1994.
- [137] Y. Fang, H. Li, Y. Ma, K. Liang, Y. Hu, S. Zhang, and H. Wang, "Dimensionality reduction of hyperspectral images based on robust spatial information using locally linear embedding," *IEEE Geoscience and Remote Sensing Letters*, vol. 11, no. 10, pp. 1712–1716, 2014.
- [138] M. Sugiyama, "Dimensionality reduction of multimodal labeled data by local fisher discriminant analysis," *Journal of machine learning research*, vol. 8, no. 5, 2007.
- [139] H.-T. Chen, H.-W. Chang, and T.-L. Liu, "Local discriminant embedding and its variants," in *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*, vol. 2. IEEE, 2005, pp. 846–853.
- [140] B.-C. Kuo and D. A. Landgrebe, "Nonparametric weighted feature extraction for classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 42, no. 5, pp. 1096–1105, 2004.
- [141] B. Kumar, O. Dikshit, A. Gupta, and M. K. Singh, "Feature extraction for hyperspectral image classification: A review," *International Journal of Remote Sensing*, 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 Transactions on Geoscience and Remote Sensing*, vol. 43, no. 3, pp. 480–491, 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 Transactions on Geoscience and Remote Sensing*, vol. 54, no. 6, pp. 3235–3247, 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 Transactions on Geoscience and Remote Sensing*, vol. 58, no. 6, pp. 3791–3808, 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 Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 11, no. 11, pp. 4141–4155, 2018.
- [146] M. Pesaresi and J. A. Benediktsson, "A new approach for the morphological segmentation of high-resolution satellite imagery," *IEEE transactions on Geoscience and Remote Sensing*, vol. 39, no. 2, pp. 309–320, 2001.
- [147] Y. Chen, X. Zhao, and X. Jia, "Spectral-spatial classification of hyperspectral data based on deep belief network," *IEEE journal of selected topics in applied earth observations and remote sensing*, vol. 8, no. 6, pp. 2381–2392, 2015.
- [148] M. E. Paoletti, J. M. Haut, J. Plaza, and A. Plaza, "Deep&dense convolutional neural network for hyperspectral image classification," *Remote Sensing*, vol. 10, no. 9, p. 1454, 2018.
- [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 Sensing*, vol. 10, no. 3, p. 395, 2018.
- [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, p. 2831, 2018.
- [151] Y. Li, W. Xie, and H. Li, "Hyperspectral image reconstruction by deep convolutional neural network for classification," *Pattern Recognition*, 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 Geoscience and Remote Sensing Letters*, vol. 14, no. 12, pp. 2365–2369, 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 Transactions on Geoscience and Remote Sensing*, vol. 57, no. 2, pp. 740–754, 2018.
- [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 Sensing*, vol. 10, no. 7, p. 1156, 2018.
- [155] Q. Liu, F. Zhou, R. Hang, and X. Yuan, "Bidirectional-convolutional lstm based spectral-spatial feature learning for hyperspectral image classification," *Remote Sensing*, vol. 9, no. 12, p. 1330, 2017.
- [156] S. K. Roy, S. Manna, T. Song, and L. Bruzzone, "Attention-based adaptive spectral-spatial kernel resnet for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 9, pp. 7831–7843, 2020.
- [157] F. Ullah, B. Zhang, R. U. Khan, T.-S. Chung, M. Attique, K. Khan, S. El Khediri, and S. Jan, "Deep edu: a deep neural collaborative filtering for educational services recommendation," *IEEE access*, vol. 8, pp. 110915–110928, 2020.
- [158] 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.
- [159] F. Ullah, B. Zhang, R. U. Khan, I. Ullah, A. Khan, and A. M. Qamar, "Visual-based items recommendation using deep neural network," in *Proceedings of the 2020 International Conference on Computing, Networks and Internet of Things*, 2020, pp. 122–126.
- [160] B. Liu, X. Yu, P. Zhang, X. Tan, A. Yu, and Z. Xue, "A semi-supervised convolutional neural network for hyperspectral image classification," *Remote Sensing Letters*, vol. 8, no. 9, pp. 839–848, 2017.
- [161] X. Kang, B. Zhuo, and P. Duan, "Semi-supervised deep learning for hyperspectral image classification," *Remote sensing letters*, vol. 10, no. 4, pp. 353–362, 2019.
- [162] Y. Wu, G. Mu, C. Qin, Q. Miao, W. Ma, and X. Zhang, "Semi-supervised hyperspectral image classification via spatial-regulated self-training," *Remote Sensing*, vol. 12, no. 1, p. 159, 2020.
- [163] Z. Zhang, "Semi-supervised hyperspectral image classification algorithm based on graph embedding and discriminative spatial information," *Microprocessors and Microsystems*, vol. 75, p. 103070, 2020.
- [164] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, 2017.
- [165] L. Liu, C. Shen, and A. Van den Hengel, "The treasure beneath convolutional layers: Cross-convolutional-layer pooling for image classification," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 4749–4757.
- [166] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*. MIT press, 2016.
- [167] 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 *International Conference on Computer Vision and Image Processing*. Springer, 2022, pp. 462–476.
- [168] T. Williams and R. Li, "Wavelet pooling for convolutional neural networks," in *International conference on learning representations*, 2018.
- [169] J. T. Springenberg, A. Dosovitskiy, T. Brox, and M. Riedmiller, "Striving for simplicity: The all convolutional net," *arXiv preprint arXiv:1412.6806*, 2014.
- [170] D. H. Hubel and T. N. Wiesel, "Receptive fields, binocular interaction and functional architecture in the cat's visual cortex," *The Journal of physiology*, vol. 160, no. 1, p. 106, 1962.
- [171] K. Fukushima, "Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position," *Biological cybernetics*, vol. 36, no. 4, pp. 193–202, 1980.
- [172] A. Voulodimos, N. Doulamis, A. Doulamis, E. Protopapadakis *et al.*, "Deep learning for computer vision: A brief review," *Computational intelligence and neuroscience*, vol. 2018, 2018.
- [173] J. Gu, Z. Wang, J. Kuen, L. Ma, A. Shahroudy, B. Shuai, T. Liu, X. Wang, G. Wang, J. Cai *et al.*, "Recent advances in convolutional neural networks," *Pattern recognition*, vol. 77, pp. 354–377, 2018.
- [174] M. Lin, Q. Chen, and S. Yan, "Network in network," *arXiv preprint arXiv:1312.4400*, 2013.
- [175] H. Gao, Y. Yang, C. Li, H. Zhou, and X. Qu, "Joint alternate small convolution and feature reuse for hyperspectral image classification," *ISPRS International Journal of Geo-Information*, vol. 7, no. 9, p. 349, 2018.
- [176] W. Zhao, L. Jiao, W. Ma, J. Zhao, J. Zhao, H. Liu, X. Cao, and S. Yang, "Superpixel-based multiple local cnn for panchromatic and multispectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 7, pp. 4141–4156, 2017.
- [177] H. Alhichri, N. Alajlan, Y. Bazi, and T. Rabczuk, "Multi-scale convolutional neural network for remote sensing scene classification," in *2018 IEEE International Conference on Electro/Information Technology (EIT)*. IEEE, 2018, pp. 1–5.
- [178] 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, p. 2644, 2017.
- [179] J. Leng, T. Li, G. Bai, Q. Dong, and H. Dong, "Cube-cnn-svm: a novel hyperspectral image classification method," in *2016 IEEE 28th International conference on tools with artificial intelligence (ICTAI)*. IEEE, 2016, pp. 1027–1034.
- [180] S. Yu, S. Jia, and C. Xu, "Convolutional neural networks for hyperspectral image classification," *Neurocomputing*, vol. 219, pp. 88–98, 2017.
- [181] H. Wu and S. Prasad, "Convolutional recurrent neural networks for hyperspectral data classification," *Remote Sensing*, vol. 9, no. 3, p. 298, 2017.
- [182] 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," *Applied Sciences*, vol. 8, no. 2, p. 212, 2018.
- [183] Q. Huang, W. Li, and X. Xie, "Convolutional neural network for medical hyperspectral image classification with kernel fusion," in *BIBE 2018: International Conference on Biological Information and Biomedical Engineering*. VDE, 2018, pp. 1–4.
- [184] K. Charmisha, V. Sowmya, and K. Soman, "Dimensionally reduced features for hyperspectral image classification using deep learning," in *ICCCE 2018: Proceedings of the International Conference on Communications and Cyber Physical Engineering 2018*. Springer, 2019, pp. 171–179.
- [185] G. Turra, S. Arrigoni, and A. Signoroni, "Cnn-based identification of hyperspectral bacterial signatures for digital microbiology," in *Image Analysis and Processing-ICIAP 2017: 19th International Conference, Catania, Italy, September 11-15, 2017, Proceedings, Part II 19*. Springer, 2017, pp. 500–510.
- [186] 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 Geoscience and Remote Sensing Letters*, vol. 15, no. 2, pp. 292–296, 2018.
- [187] J. M. Haut, M. E. Paoletti, J. Plaza, A. Plaza, and J. Li, "Hyperspectral image classification using random occlusion data augmentation," *IEEE Geoscience and Remote Sensing Letters*, vol. 16, no. 11, pp. 1751–1755, 2019.
- [188] Y. Xu, B. Du, F. Zhang, and L. Zhang, "Hyperspectral image classification via a random patches network," *ISPRS journal of photogrammetry and remote sensing*, vol. 142, pp. 344–357, 2018.
- [189] Y. Wang, T. Song, Y. Xie, and S. K. Roy, "A probabilistic neighbourhood pooling-based attention network for hyperspectral image classification," *Remote Sensing Letters*, vol. 13, no. 1, pp. 65–75, 2022.
- [190] 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 Sensing*, vol. 9, no. 6, p. 618, 2017.
- [191] Y. Chen, L. Zhu, P. Ghamisi, X. Jia, G. Li, and L. Tang, "Hyperspectral images classification with gabor filtering and convolutional neural network," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 12, pp. 2355–2359, 2017.
- [192] J. Zhu, L. Fang, and P. Ghamisi, "Deformable convolutional neural networks for hyperspectral image classification," *IEEE Geoscience and Remote Sensing Letters*, vol. 15, no. 8, pp. 1254–1258, 2018.
- [193] L. Ran, Y. Zhang, W. Wei, and Q. Zhang, "A hyperspectral image classification framework with spatial pixel pair features," *Sensors*, vol. 17, no. 10, p. 2421, 2017.
- [194] Z. Zhong, J. Li, Z. Luo, and M. Chapman, "Spectral-spatial residual network for hyperspectral image classification: A 3-d deep learning framework," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 2, pp. 847–858, 2017.
- [195] M. E. Paoletti, J. M. Haut, J. Plaza, and A. Plaza, "A new deep convolutional neural network for fast hyperspectral image classification," *ISPRS journal of photogrammetry and remote sensing*, vol. 145, pp. 120–147, 2018.
- [196] S. Li, X. Zhu, Y. Liu, and J. Bao, "Adaptive spatial-spectral feature learning for hyperspectral image classification," *IEEE access*, vol. 7, pp. 61 534–61 547, 2019.

- [197] 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 *2021 11th Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing (WHISPERS)*. IEEE, 2021, pp. 1–5.
- [198] 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. 179 575–179 591, 2020.
- [199] 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 Transactions on Geoscience and Remote Sensing*, vol. 57, no. 8, pp. 5813–5828, 2019.
- [200] S. Jia, Z. Lin, M. Xu, Q. Huang, J. Zhou, X. Jia, and Q. Li, "A lightweight convolutional neural network for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 5, pp. 4150–4163, 2020.
- [201] 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 Transactions on Geoscience and Remote Sensing*, vol. 58, no. 8, pp. 5277–5290, 2020.
- [202] S. K. Roy, R. Mondal, M. E. Paoletti, J. M. Haut, and A. Plaza, "Morphological convolutional neural networks for hyperspectral image classification," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 8689–8702, 2021.
- [203] Y. Li, H. Zhang, and Q. Shen, "Spectral-spatial classification of hyperspectral imagery with 3d convolutional neural network," *Remote Sensing*, vol. 9, no. 1, p. 67, 2017.
- [204] S. K. Roy, S. R. Dubey, S. Chatterjee, and B. Baran Chaudhuri, "Fusenet: fused squeeze-and-excitation network for spectral-spatial hyperspectral image classification," *IET Image Processing*, vol. 14, no. 8, pp. 1653–1661, 2020.
- [205] 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 Transactions on Geoscience and Remote Sensing*, vol. 55, no. 10, pp. 5585–5599, 2017.
- [206] H. Zhang, Y. Li, Y. Zhang, and Q. Shen, "Spectral-spatial classification of hyperspectral imagery using a dual-channel convolutional neural network," *Remote sensing letters*, vol. 8, no. 5, pp. 438–447, 2017.
- [207] M. He, B. Li, and H. Chen, "Multi-scale 3d deep convolutional neural network for hyperspectral image classification," in *2017 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2017, pp. 3904–3908.
- [208] H. Dong, L. Zhang, and B. Zou, "Band attention convolutional networks for hyperspectral image classification," *arXiv preprint arXiv:1906.04379*, 2019.
- [209] N. He, M. E. Paoletti, J. M. Haut, L. Fang, S. Li, A. Plaza, and J. Plaza, "Feature extraction with multiscale covariance maps for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 2, pp. 755–769, 2018.
- [210] G. Cheng, Z. Li, J. Han, X. Yao, and L. Guo, "Exploring hierarchical convolutional features for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 11, pp. 6712–6722, 2018.
- [211] Z. Gong, P. Zhong, Y. Yu, W. Hu, and S. Li, "A cnn with multiscale convolution and diversified metric for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 6, pp. 3599–3618, 2019.
- [212] P. Zhong, N. Peng, and R. Wang, "Learning to diversify patch-based priors for remote sensing image restoration," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 8, no. 11, pp. 5225–5245, 2015.
- [213] L. Liu, Z. Shi, B. Pan, N. Zhang, H. Luo, and X. Lan, "Multiscale deep spatial feature extraction using virtual rgb image for hyperspectral imagery classification," *Remote sensing*, vol. 12, no. 2, p. 280, 2020.
- [214] X. Ma, A. Fu, J. Wang, H. Wang, and B. Yin, "Hyperspectral image classification based on deep deconvolution network with skip architecture," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 8, pp. 4781–4791, 2018.
- [215] 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 Systems with Applications*, vol. 129, pp. 246–259, 2019.
- [216] S. K. Roy, S. Das, T. Song, and B. Chanda, "Darecnet-bs: Unsupervised dual-attention reconstruction network for hyperspectral band selection," *IEEE Geoscience and Remote Sensing Letters*, vol. 18, no. 12, pp. 2152–2156, 2020.
- [217] 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 Transactions on Geoscience and Remote Sensing*, vol. 57, no. 9, pp. 6808–6820, 2019.
- [218] 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 Geoscience and Remote Sensing Letters*, vol. 19, pp. 1–5, 2021.
- [219] W. Song, S. Li, L. Fang, and T. Lu, "Hyperspectral image classification with deep feature fusion network," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 6, pp. 3173–3184, 2018.
- [220] 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 Geoscience and Remote Sensing Letters*, vol. 17, no. 2, pp. 277–281, 2019.
- [221] B. Zhang, S. Li, X. Jia, L. Gao, and M. Peng, "Adaptive markov random field approach for classification of hyperspectral imagery," *IEEE Geoscience and Remote Sensing Letters*, vol. 8, no. 5, pp. 973–977, 2011.
- [222] 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 Geoscience and Remote Sensing Letters*, vol. 19, pp. 1–5, 2021.
- [223] 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 Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–19, 2021.
- [224] 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 Transactions on Geoscience and Remote Sensing*, 2023.
- [225] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," *arXiv preprint arXiv:1609.02907*, 2016.
- [226] D. Hong, L. Gao, J. Yao, B. Zhang, A. Plaza, and J. Chanussot, "Graph convolutional networks for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 7, pp. 5966–5978, 2020.
- [227] 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 Geoscience and Remote Sensing Letters*, vol. 16, no. 2, pp. 241–245, 2018.
- [228] S. Wan, C. Gong, P. Zhong, S. Pan, G. Li, and J. Yang, "Hyperspectral image classification with context-aware dynamic graph convolutional network," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 1, pp. 597–612, 2020.
- [229] J. Zhu, L. Wu, H. Hao, X. Song, and Y. Lu, "Auto-encoder based for high spectral dimensional data classification and visualization," in *2017 IEEE Second International Conference on Data Science in Cyberspace (DSC)*. IEEE, 2017, pp. 350–354.
- [230] A. Hassanzadeh, A. Kaarna, and T. Kauranne, "Unsupervised multi-manifold classification of hyperspectral remote sensing images with contractive autoencoder," in *Image Analysis: 20th Scandinavian Conference, SCIA 2017, Tromsø, Norway, June 12–14, 2017, Proceedings, Part II 20*. Springer, 2017, pp. 169–180.
- [231] Y. Wang, Y. Jiang, Y. Wu, and Z.-H. Zhou, "Multi-manifold clustering," in *PRICAI 2010: Trends in Artificial Intelligence: 11th Pacific Rim International Conference on Artificial Intelligence, Daegu, Korea, August 30–September 2, 2010. Proceedings 11*. Springer, 2010, pp. 280–291.
- [232] S. Rifai, P. Vincent, X. Muller, X. Glorot, and Y. Bengio, "Contractive auto-encoders: Explicit invariance during feature extraction," in *Proceedings of the 28th international conference on international conference on machine learning*, 2011, pp. 833–840.
- [233] X. Zhang, Y. Liang, C. Li, N. Huyen, L. Jiao, and H. Zhou, "Recursive autoencoders-based unsupervised feature learning for hyperspectral image classification," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 11, pp. 1928–1932, 2017.
- [234] S. Hao, W. Wang, Y. Ye, T. Nie, and L. Bruzzone, "Two-stream deep architecture for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 4, pp. 2349–2361, 2017.
- [235] K. He, J. Sun, and X. Tang, "Guided image filtering," *IEEE transactions on pattern analysis and machine intelligence*, vol. 35, no. 6, pp. 1397–1409, 2012.
- [236] X. Sun, F. Zhou, J. Dong, F. Gao, Q. Mu, and X. Wang, "Encoding spectral and spatial context information for hyperspectral image classification," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 12, pp. 2250–2254, 2017.

- [237] 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," *European journal of remote sensing*, vol. 50, no. 1, pp. 47–63, 2017.
- [238] 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 Physics & Technology*, vol. 86, pp. 77–89, 2017.
- [239] 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.
- [240] M. Ahmad, A. M. Khan, M. Mazzara, and S. Distefano, "Multi-layer extreme learning machine-based autoencoder for hyperspectral image classification," in *VISIGRAPP (4: VISAPP)*, 2019, pp. 75–82.
- [241] P. Zhou, J. Han, G. Cheng, and B. Zhang, "Learning compact and discriminative stacked autoencoder for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 7, pp. 4823–4833, 2019.
- [242] R. Lan, Z. Li, Z. Liu, T. Gu, and X. Luo, "Hyperspectral image classification using k-sparse denoising autoencoder and spectral-restricted spatial characteristics," *Applied Soft Computing*, vol. 74, pp. 693–708, 2019.
- [243] S. Paul and D. N. Kumar, "Spectral-spatial classification of hyperspectral data with mutual information based segmented stacked autoencoder approach," *ISPRS journal of photogrammetry and remote sensing*, vol. 138, pp. 265–280, 2018.
- [244] 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," *Journal of the Indian Society of Remote Sensing*, vol. 47, pp. 1989–1997, 2019.
- [245] G. E. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," *Neural computation*, vol. 18, no. 7, pp. 1527–1554, 2006.
- [246] N. Zhang, S. Ding, J. Zhang, and Y. Xue, "An overview on restricted boltzmann machines," *Neurocomputing*, vol. 275, pp. 1186–1199, 2018.
- [247] B. Ayhan and C. Kwan, "Application of deep belief network to land cover classification using hyperspectral images," in *Advances in Neural Networks-ISNN 2017: 14th International Symposium, ISNN 2017, Sapporo, Hakodate, and Muroran, Hokkaido, Japan, June 21–26, 2017, Proceedings, Part I 14*. Springer, 2017, pp. 269–276.
- [248] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov, "Improving neural networks by preventing co-adaptation of feature detectors," *arXiv preprint arXiv:1207.0580*, 2012.
- [249] U. Shaham, X. Cheng, O. Dror, A. Jaffe, B. Nadler, J. Chang, and Y. Kluger, "A deep learning approach to unsupervised ensemble learning," in *International conference on machine learning*. PMLR, 2016, pp. 30–39.
- [250] H. Xiong, A. J. Rodríguez-Sánchez, S. Szedmak, and J. Piater, "Diversity priors for learning early visual features," *Frontiers in computational neuroscience*, vol. 9, p. 104, 2015.
- [251] P. Zhong, Z. Gong, S. Li, and C.-B. Schönlieb, "Learning to diversify deep belief networks for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 6, pp. 3516–3530, 2017.
- [252] J. Li, B. Xi, Y. Li, Q. Du, and K. Wang, "Hyperspectral classification based on texture feature enhancement and deep belief networks," *Remote Sensing*, vol. 10, no. 3, p. 396, 2018.
- [253] 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 Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 12, no. 2, pp. 627–636, 2019.
- [254] S. Li, W. Song, L. Fang, Y. Chen, P. Ghamisi, and J. A. Benediktsson, "Deep learning for hyperspectral image classification: An overview," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 9, pp. 6690–6709, 2019.
- [255] A. Sellami and I. Farah, "Spectra-spatial graph-based deep restricted boltzmann networks for hyperspectral image classification," in *2019 Photonics & Electromagnetics Research Symposium-Spring (PIERS-Spring)*. IEEE, 2019, pp. 1055–1062.
- [256] R. J. Williams and D. Zipser, "A learning algorithm for continually running fully recurrent neural networks," *Neural computation*, vol. 1, no. 2, pp. 270–280, 1989.
- [257] M. E. Paoletti, J. M. Haut, J. Plaza, and A. Plaza, "Scalable recurrent neural network for hyperspectral image classification," *The Journal of Supercomputing*, vol. 76, pp. 8866–8882, 2020.
- [258] F. Zhou, R. Hang, Q. Liu, and X. Yuan, "Hyperspectral image classification using spectral-spatial lstms," *Neurocomputing*, vol. 328, pp. 39–47, 2019.
- [259] 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 Networks*, vol. 105, pp. 346–355, 2018.
- [260] H. Wu and S. Prasad, "Semi-supervised deep learning using pseudo labels for hyperspectral image classification," *IEEE Transactions on Image Processing*, vol. 27, no. 3, pp. 1259–1270, 2017.
- [261] F. Zhou, R. Hang, Q. Liu, and X. Yuan, "Integrating convolutional neural network and gated recurrent unit for hyperspectral image spectral-spatial classification," in *Pattern Recognition and Computer Vision: First Chinese Conference, PRCV 2018, Guangzhou, China, November 23–26, 2018, Proceedings, Part IV 1*. Springer, 2018, pp. 409–420.
- [262] H. Luo, "Shorten spatial-spectral rnn with parallel-gru for hyperspectral image classification," *arXiv preprint arXiv:1810.12563*, 2018.
- [263] —, "Shorten spatial-spectral rnn with parallel-gru for hyperspectral image classification," *arXiv preprint arXiv:1810.12563*, 2018.
- [264] X. Yang, Y. Ye, X. Li, R. Y. Lau, X. Zhang, and X. Huang, "Hyperspectral image classification with deep learning models," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 9, pp. 5408–5423, 2018.
- [265] M. Seydgar, A. Alizadeh Naeini, M. Zhang, W. Li, and M. Satri, "3-d convolution-recurrent networks for spectral-spatial classification of hyperspectral images," *Remote Sensing*, vol. 11, no. 7, p. 883, 2019.
- [266] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *Journal of big data*, vol. 6, no. 1, pp. 1–48, 2019.
- [267] 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.
- [268] X. Yu, X. Wu, C. Luo, and P. Ren, "Deep learning in remote sensing scene classification: a data augmentation enhanced convolutional neural network framework," *GIScience & Remote Sensing*, vol. 54, no. 5, pp. 741–758, 2017.
- [269] W. Li, C. Chen, M. Zhang, H. Li, and Q. Du, "Data augmentation for hyperspectral image classification with deep cnn," *IEEE Geoscience and Remote Sensing Letters*, vol. 16, no. 4, pp. 593–597, 2018.
- [270] X. Cao, J. Yao, Z. Xu, and D. Meng, "Hyperspectral image classification with convolutional neural network and active learning," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 7, pp. 4604–4616, 2020.
- [271] 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 *2019 9th International Conference on Information Science and Technology (ICIST)*. IEEE, 2019, pp. 362–367.
- [272] J. Nalepa, M. Myller, and M. Kawulok, "Training-and test-time data augmentation for hyperspectral image segmentation," *IEEE Geoscience and Remote Sensing Letters*, vol. 17, no. 2, pp. 292–296, 2019.
- [273] —, "Hyperspectral data augmentation," *arXiv preprint arXiv:1903.05580*, 2019.
- [274] J. E. Van Engelen and H. H. Hoos, "A survey on semi-supervised learning," *Machine learning*, vol. 109, no. 2, pp. 373–440, 2020.
- [275] N. N. Pise and P. Kulkarni, "A survey of semi-supervised learning methods," in *2008 International conference on computational intelligence and security*, vol. 2. IEEE, 2008, pp. 30–34.
- [276] S. S. Sawant and M. Prabukumar, "Semi-supervised techniques based hyper-spectral image classification: a survey," *2017 Innovations in Power and Advanced Computing Technologies (i-PACT)*, pp. 1–8, 2017.
- [277] 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 Sensing*, vol. 10, no. 4, p. 574, 2018.
- [278] S. Zhou, Z. Xue, and P. Du, "Semisupervised stacked autoencoder with cotraining for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 6, pp. 3813–3826, 2019.
- [279] 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 Transactions on Geoscience and Remote Sensing*, vol. 56, no. 1, pp. 3–16, 2017.
- [280] M. S. Aydemir and G. Bilgin, "Semisupervised hyperspectral image classification using small sample sizes," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 5, pp. 621–625, 2017.

- [281] P. Bateni, R. Goyal, V. Masrani, F. Wood, and L. Sigal, "Improved few-shot visual classification," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 14 493–14 502.
- [282] B. Liu, X. Yu, A. Yu, P. Zhang, G. Wan, and R. Wang, "Deep few-shot learning for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 4, pp. 2290–2304, 2018.
- [283] 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 Transactions on Neural Networks and Learning Systems*, 2022.
- [284] 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 Sensing*, vol. 14, no. 9, p. 2246, 2022.
- [285] Y. Xu, Y. Zhang, T. Yue, C. Yu, and H. Li, "Graph-based domain adaptation few-shot learning for hyperspectral image classification," *Remote Sensing*, vol. 15, no. 4, p. 1125, 2023.
- [286] H. Wang, X. Wang, and Y. Cheng, "Graph meta transfer network for heterogeneous few-shot hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–12, 2023.
- [287] Q. Ran, Y. Zhou, D. Hong, M. Bi, L. Ni, X. Li, and M. Ahmad, "Deep transformer and few-shot learning for hyperspectral image classification," *CAAI Transactions on Intelligence Technology*, 2023.
- [288] Y. Peng, Y. Liu, B. Tu, and Y. Zhang, "Convolutional transformer-based few-shot learning for cross-domain hyperspectral image classification," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 16, pp. 1335–1349, 2023.
- [289] Z. Huang, H. Tang, Y. Li, and W. Xie, "Hfc-sst: improved spatial-spectral transformer for hyperspectral few-shot classification," *Journal of Applied Remote Sensing*, vol. 17, no. 2, pp. 026 509–026 509, 2023.
- [290] A. Creswell, T. White, V. Dumoulin, K. Arulkumaran, B. Sengupta, and A. A. Bharath, "Generative adversarial networks: An overview," *IEEE signal processing magazine*, vol. 35, no. 1, pp. 53–65, 2018.
- [291] Y. Zhan, D. Hu, Y. Wang, and X. Yu, "Semisupervised hyperspectral image classification based on generative adversarial networks," *IEEE Geoscience and Remote Sensing Letters*, vol. 15, no. 2, pp. 212–216, 2017.
- [292] Z. He, H. Liu, Y. Wang, and J. Hu, "Generative adversarial networks-based semi-supervised learning for hyperspectral image classification," *Remote Sensing*, vol. 9, no. 10, p. 1042, 2017.
- [293] L. Zhu, Y. Chen, P. Ghamisi, and J. A. Benediktsson, "Generative adversarial networks for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 9, pp. 5046–5063, 2018.
- [294] Y. Zhan, K. Wu, W. Liu, J. Qin, Z. Yang, Y. Medjadba, G. Wang, and X. Yu, "Semi-supervised classification of hyperspectral data based on generative adversarial networks and neighborhood majority voting," in *IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, 2018, pp. 5756–5759.
- [295] 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 Transactions on Geoscience and Remote Sensing*, vol. 57, no. 8, pp. 5329–5343, 2019.
- [296] Z. Zhong, J. Li, D. A. Clausi, and A. Wong, "Generative adversarial networks and conditional random fields for hyperspectral image classification," *IEEE transactions on cybernetics*, vol. 50, no. 7, pp. 3318–3329, 2019.
- [297] X. Wang, K. Tan, Q. Du, Y. Chen, and P. Du, "Caps-tripleGAN: Gan-assisted capsnet for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 9, pp. 7232–7245, 2019.
- [298] Z. Xue, "Semi-supervised convolutional generative adversarial network for hyperspectral image classification," *IET Image Processing*, vol. 14, no. 4, pp. 709–719, 2020.
- [299] 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 Geoscience and Remote Sensing Letters*, vol. 18, no. 3, pp. 523–527, 2020.
- [300] T. Alipour-Fard and H. Arefi, "Structure aware generative adversarial networks for hyperspectral image classification," *Ieee Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 13, pp. 5424–5438, 2020.
- [301] 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 Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–15, 2021.
- [302] D. Hong, B. Zhang, H. Li, Y. Li, J. Yao, C. Li, M. Werner, J. Chanussot, A. Zipf, and X. X. Zhu, "Cross-city matters: A multimodal remote sensing benchmark dataset for cross-city semantic segmentation using high-resolution domain adaptation networks," *Remote Sensing of Environment*, vol. 299, p. 113856, 2023.
- [303] J. Yang, Y.-Q. Zhao, and J. C.-W. Chan, "Learning and transferring deep joint spectral-spatial features for hyperspectral classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 8, pp. 4729–4742, 2017.
- [304] L. Windrim, A. Melkumyan, R. J. Murphy, A. Chlingaryan, and R. Ramakrishnan, "Pretraining for hyperspectral convolutional neural network classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 5, pp. 2798–2810, 2018.
- [305] 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 Sensing*, vol. 10, no. 9, p. 1425, 2018.
- [306] O. Day and T. M. Khoshgoftaar, "A survey on heterogeneous transfer learning," *Journal of Big Data*, vol. 4, pp. 1–42, 2017.
- [307] J. Lin, R. Ward, and Z. J. Wang, "Deep transfer learning for hyperspectral image classification," in *2018 IEEE 20th International Workshop on Multimedia Signal Processing (MMSP)*. IEEE, 2018, pp. 1–5.
- [308] X. Li, L. Zhang, B. Du, L. Zhang, and Q. Shi, "Iterative reweighting heterogeneous transfer learning framework for supervised remote sensing image classification," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 10, no. 5, pp. 2022–2035, 2017.
- [309] Y. Liu and C. Xiao, "Transfer learning for hyperspectral image classification using convolutional neural network," in *MIPPR 2019: Remote Sensing Image Processing, Geographic Information Systems, and Other Applications*, vol. 11432. SPIE, 2020, pp. 79–84.
- [310] J. Lin, C. He, Z. J. Wang, and S. Li, "Structure preserving transfer learning for unsupervised hyperspectral image classification," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 10, pp. 1656–1660, 2017.
- [311] R. Pires de Lima and K. Marfurt, "Convolutional neural network for remote-sensing scene classification: Transfer learning analysis," *Remote Sensing*, vol. 12, no. 1, p. 86, 2019.
- [312] D. Hong, B. Zhang, X. Li, Y. Li, C. Li, J. Yao, N. Yokoya, H. Li, X. Jia, A. Plaza *et al.*, "Spectralgpt: Spectral foundation model," *arXiv preprint arXiv:2311.07113*, 2023.
- [313] X. He, Y. Chen, and Z. Lin, "Spatial-spectral transformer for hyperspectral image classification," *Remote Sensing*, vol. 13, no. 3, p. 498, 2021.
- [314] Y. Qing, W. Liu, L. Feng, and W. Gao, "Improved transformer net for hyperspectral image classification," *Remote Sensing*, vol. 13, no. 11, p. 2216, 2021.
- [315] L. Sun, G. Zhao, Y. Zheng, and Z. Wu, "Spectral-spatial feature tokenization transformer for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–14, 2022.
- [316] B. Liu, A. Yu, K. Gao, X. Tan, Y. Sun, and X. Yu, "Dss-trm: Deep spatial-spectral transformer for hyperspectral image classification," *European Journal of Remote Sensing*, vol. 55, no. 1, pp. 103–114, 2022.
- [317] 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 Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–13, 2022.
- [318] L. Yang, Y. Yang, J. Yang, N. Zhao, L. Wu, L. Wang, and T. Wang, "Fusionnet: a convolution-transformer fusion network for hyperspectral image classification," *Remote Sensing*, vol. 14, no. 16, p. 4066, 2022.
- [319] J. Zhang, Z. Meng, F. Zhao, H. Liu, and Z. Chang, "Convolution transformer mixer for hyperspectral image classification," *IEEE Geoscience and Remote Sensing Letters*, vol. 19, pp. 1–5, 2022.
- [320] C. Zhao, B. Qin, S. Feng, W. Zhu, W. Sun, W. Li, and X. Jia, "Hyperspectral image classification with multi-attention transformer and adaptive superpixel segmentation-based active learning," *IEEE Transactions on Image Processing*, 2023.
- [321] Z. Qiu, J. Xu, J. Peng, and W. Sun, "Cross-channel dynamic spatial-spectral fusion transformer for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, 2023.
- [322] 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 Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–15, 2023.

- [323] S. Hao, Y. Xia, and Y. Ye, "Generative adversarial network with transformer for hyperspectral image classification," *IEEE Geoscience and Remote Sensing Letters*, 2023.
- [324] X. Cao, H. Lin, S. Guo, T. Xiong, and L. Jiao, "Transformer-based masked autoencoder with contrastive loss for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, 2023.
- [325] L. Liang, Y. Zhang, S. Zhang, J. Li, A. Plaza, and X. Kang, "Fast hyperspectral image classification combining transformers and simam-based cnns," *IEEE Transactions on Geoscience and Remote Sensing*, 2023.
- [326] R. Ganti and A. Gray, "Upal: Unbiased pool based active learning," in *Artificial Intelligence and Statistics*. PMLR, 2012, pp. 422–431.
- [327] P. Melville and R. J. Mooney, "Diverse ensembles for active learning," in *Proceedings of the twenty-first international conference on Machine learning*, 2004, p. 74.
- [328] C. C. Aggarwal, X. Kong, Q. Gu, J. Han, and S. Y. Philip, "Active learning: A survey," in *Data classification*. Chapman and Hall/CRC, 2014, pp. 599–634.
- [329] H. S. Seung, M. Opper, and H. Sompolinsky, "Query by committee," in *Proceedings of the fifth annual workshop on Computational learning theory*, 1992, pp. 287–294.
- [330] B. Settles, "Active learning literature survey," 2009.
- [331] C. Liu, L. He, Z. Li, and J. Li, "Feature-driven active learning for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 1, pp. 341–354, 2017.
- [332] Y. Zhang, G. Cao, X. Li, B. Wang, and P. Fu, "Active semi-supervised random forest for hyperspectral image classification," *Remote Sensing*, vol. 11, no. 24, p. 2974, 2019.
- [333] 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 Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 10, no. 1, pp. 347–359, 2016.
- [334] Z. Xue, S. Zhou, and P. Zhao, "Active learning improved by neighborhoods and superpixels for hyperspectral image classification," *IEEE Geoscience and Remote Sensing Letters*, vol. 15, no. 3, pp. 469–473, 2018.
- [335] K. Bhardwaj, A. Das, and S. Patra, "Spectral-spatial active learning with attribute profile for hyperspectral image classification," in *International Conference on Intelligent Computing and Smart Communication 2019: Proceedings of ICSC 2019*. Springer, 2020, pp. 1219–1229.
- [336] S. Patra, K. Bhardwaj, and L. Bruzzone, "A spectral-spatial multicriteria active learning technique for hyperspectral image classification," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 10, no. 12, pp. 5213–5227, 2017.
- [337] Z. Zhang and M. M. Crawford, "A batch-mode regularized multimetric active learning framework for classification of hyperspectral images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 11, pp. 6594–6609, 2017.
- [338] X. Xu, J. Li, and S. Li, "Multiview intensity-based active learning for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 2, pp. 669–680, 2017.
- [339] M. K. Pradhan, S. Minz, and V. K. Shrivastava, "Fisher discriminant ratio based multiview active learning for the classification of remote sensing images," in *2018 4th International Conference on Recent Advances in Information Technology (RAIT)*. IEEE, 2018, pp. 1–6.
- [340] Z. Zhang, E. Pasolli, and M. M. Crawford, "An adaptive multiview active learning approach for spectral-spatial classification of hyperspectral images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 4, pp. 2557–2570, 2019.
- [341] Y. Li, T. Lu, and S. Li, "Subpixel-pixel-superpixel-based multiview active learning for hyperspectral images classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 7, pp. 4976–4988, 2020.
- [342] P. Liu, H. Zhang, and K. B. Eom, "Active deep learning for classification of hyperspectral images," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 10, no. 2, pp. 712–724, 2016.
- [343] Y. Sun, J. Li, W. Wang, A. Plaza, and Z. Chen, "Active learning based autoencoder for hyperspectral imagery classification," in *2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*. IEEE, 2016, pp. 469–472.
- [344] 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 Transactions on Geoscience and Remote Sensing*, vol. 56, no. 11, pp. 6440–6461, 2018.
- [345] J. Lin, L. Zhao, S. Li, R. Ward, and Z. J. Wang, "Active-learning-incorporated deep transfer learning for hyperspectral image classification," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 11, no. 11, pp. 4048–4062, 2018.
- [346] 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 Transactions on Geoscience and Remote Sensing*, vol. 57, no. 3, pp. 1741–1754, 2018.
- [347] C. Deng, X. Liu, C. Li, and D. Tao, "Active multi-kernel domain adaptation for hyperspectral image classification," *Pattern Recognition*, vol. 77, pp. 306–315, 2018.
- [348] M. Ahmad, M. Mazzara, R. A. Raza, S. Distefano, M. Asif, M. S. Sarfraz, A. M. Khan, and A. Sohaib, "Multiclass non-randomized spectral-spatial active learning for hyperspectral image classification," *Applied Sciences*, vol. 10, no. 14, p. 4739, 2020.
- [349] J. Yao, X. Cao, D. Hong, X. Wu, D. Meng, J. Chanussot, and Z. Xu, "Semi-active convolutional neural networks for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–15, 2022.
- [350] R. E. Schapire, "The boosting approach to machine learning: An overview," *Nonlinear estimation and classification*, pp. 149–171, 2003.
- [351] —, "Explaining adaboost," *Empirical Inference: Festschrift in Honor of Vladimir N. Vapnik*, pp. 37–52, 2013.
- [352] G. Biau and E. Scornet, "A random forest guided tour," *Test*, vol. 25, pp. 197–227, 2016.
- [353] 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 Transactions on Geoscience and Remote Sensing*, vol. 48, no. 7, pp. 2880–2889, 2010.
- [354] W. Yao, C. Lian, and L. Bruzzone, "A cnn ensemble based on a spectral feature refining module for hyperspectral image classification," *Remote Sensing*, vol. 14, no. 19, p. 4982, 2022.
- [355] J. Nalepa, M. Myller, L. Tulczyjew, and M. Kawulok, "Deep ensembles for hyperspectral image data classification and unmixing," *Remote Sensing*, vol. 13, no. 20, p. 4133, 2021.
- [356] 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 Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–15, 2022.
- [357] Y. Chen, Y. Wang, Y. Gu, X. He, P. Ghamisi, and X. Jia, "Deep learning ensemble for hyperspectral image classification," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 12, no. 6, pp. 1882–1897, 2019.
- [358] H. Su, Y. Hu, H. Lu, W. Sun, and Q. Du, "Diversity-driven multikernel collaborative representation ensemble for hyperspectral image classification," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 15, pp. 2861–2876, 2022.
- [359] H. Lu, H. Su, J. Hu, and Q. Du, "Dynamic ensemble learning with multi-view kernel collaborative subspace clustering for hyperspectral image classification," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 15, pp. 2681–2695, 2022.
- [360] 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 Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 3988–3999, 2021.
- [361] Q. Lv, W. Feng, Y. Quan, Q. Li, G. Dauphin, L. Gao, G. Zhao, and M. Xing, "Ensemble cnn with enhanced feature subspaces for imbalanced hyperspectral image classification," in *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*. IEEE, 2021, pp. 3669–3672.
- [362] Q. Li, B. Zheng, B. Tu, J. Wang, and C. Zhou, "Ensemble emd-based spectral-spatial feature extraction for hyperspectral image classification," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 13, pp. 5134–5148, 2020.
- [363] H. Su, Y. Yu, Q. Du, and P. Du, "Ensemble learning for hyperspectral image classification using tangent collaborative representation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 6, pp. 3778–3790, 2020.
- [364] W. Feng, W. Huang, G. Dauphin, J. Xia, Y. Quan, H. Ye, and Y. Dong, "Ensemble margin based semi-supervised random forest for the classification of hyperspectral image with limited training data," in *IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, 2019, pp. 971–974.
- [365] 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 Sensing*, vol. 14, no. 17, p. 4236, 2022.

- [366] —, “Es2fl: Ensemble self-supervised feature learning for small sample classification of hyperspectral images,” *Remote Sensing*, vol. 14, no. 17, p. 4236, 2022.
- [367] 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 Geoscience and Remote Sensing Letters*, vol. 19, pp. 1–5, 2022.
- [368] X. Yin, R. Wang, X. Liu, and Y. Cai, “Deep forest-based classification of hyperspectral images,” in *2018 37th Chinese Control Conference (CCC)*. IEEE, 2018, pp. 10367–10372.
- [369] 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 and Applications*, vol. 78, pp. 3411–3433, 2019.
- [370] X. Liu, R. Wang, Z. Cai, Y. Cai, and X. Yin, “Deep multigrained cascade forest for hyperspectral image classification,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 10, pp. 8169–8183, 2019.
- [371] X. Cao, L. Wen, Y. Ge, J. Zhao, and L. Jiao, “Rotation-based deep forest for hyperspectral imagery classification,” *IEEE Geoscience and Remote Sensing Letters*, vol. 16, no. 7, pp. 1105–1109, 2019.
- [372] X. Cao, R. Li, Y. Ge, B. Wu, and L. Jiao, “Densely connected deep random forest for hyperspectral imagery classification,” *International journal of remote sensing*, vol. 40, no. 9, pp. 3606–3622, 2019.
- [373] B. Liu, W. Guo, X. Chen, K. Gao, X. Zuo, R. Wang, and A. Yu, “Morphological attribute profile cube and deep random forest for small sample classification of hyperspectral image,” *IEEE Access*, vol. 8, pp. 117096–117108, 2020.
- [374] J. Shi, T. Shao, X. Liu, X. Zhang, Z. Zhang, and Y. Lei, “Evolutionary multitask ensemble learning model for hyperspectral image classification,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 936–950, 2020.
- [375] X. Liu, Q. Hu, Y. Cai, and Z. Cai, “Extreme learning machine-based ensemble transfer learning for hyperspectral image classification,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 13, pp. 3892–3902, 2020.
- [376] 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 Geoscience and Remote Sensing Letters*, vol. 16, no. 12, pp. 1879–1883, 2019.
- [377] Y. Zhang, G. Cao, A. Shafique, and P. Fu, “Label propagation ensemble for hyperspectral image classification,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 12, no. 9, pp. 3623–3636, 2019.
- [378] L. Fang, W. Zhao, N. He, and J. Zhu, “Multiscale cnns ensemble based self-learning for hyperspectral image classification,” *IEEE Geoscience and Remote Sensing Letters*, vol. 17, no. 9, pp. 1593–1597, 2020.
- [379] Y. Zhang, G. Cao, and X. Li, “Multiview-based random rotation ensemble pruning for hyperspectral image classification,” *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–14, 2020.
- [380] W. Feng, Y. Quan, G. Dauphin, Q. Li, L. Gao, W. Huang, J. Xia, W. Zhu, and M. Xing, “Semi-supervised rotation forest based on ensemble margin theory for the classification of hyperspectral image with limited training data,” *Information Sciences*, vol. 575, pp. 611–638, 2021.
- [381] X. He and Y. Chen, “Transferring cnn ensemble for hyperspectral image classification,” *IEEE Geoscience and Remote Sensing Letters*, vol. 18, no. 5, pp. 876–880, 2020.
- [382] W. Feng, Y. Quan, G. Dauphin, X. Zhong, Q. Li, M. Xing, and W. Huang, “Two-step ensemble based class noise cleaning method for hyperspectral image classification,” in *IGARSS 2020-2020 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, 2020, pp. 830–833.
- [383] H. Lu, H. Su, P. Zheng, Y. Gao, and Q. Du, “Weighted residual dynamic ensemble learning for hyperspectral image classification,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 15, pp. 6912–6927, 2022.
- [384] J. Yao, B. Zhang, C. Li, D. Hong, and J. Chanussot, “Extended vision transformer (exvit) for land use and land cover classification: A multi-modal deep learning framework,” *IEEE Transactions on Geoscience and Remote Sensing*, 2023.
- [385] J. Yao, D. Hong, H. Wang, H. Liu, and J. Chanussot, “Ucsl: Towards unsupervised common subspace learning for cross-modal image classification,” *IEEE Transactions on Geoscience and Remote Sensing*, 2023.
- [386] D. Hong, J. Yao, D. Meng, Z. Xu, and J. Chanussot, “Multimodal gans: Toward crossmodal hyperspectral–multispectral image segmentation,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 6, pp. 5103–5113, 2020.
- [387] 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 Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 16, pp. 3635–3644, 2022.
- [388] R. O. Green, M. L. Eastwood, C. M. Sarture, T. G. Chrien, M. Aronsson, B. J. Chippendale, J. A. Faust, B. E. Pavri, C. J. Chovit, M. Solis *et al.*, “Imaging spectroscopy and the airborne visible/infrared imaging spectrometer (aviris),” *Remote sensing of environment*, vol. 65, no. 3, pp. 227–248, 1998.
- [389] X. Huang and L. Zhang, “A comparative study of spatial approaches for urban mapping using hyperspectral rosis images over pavia city, northern italy,” *International Journal of Remote Sensing*, vol. 30, no. 12, pp. 3205–3221, 2009.
- [390] X. Xu, J. Li, and A. Plaza, “Fusion of hyperspectral and lidar data using morphological component analysis,” in *2016 IEEE international geoscience and remote sensing symposium (IGARSS)*. IEEE, 2016, pp. 3575–3578.
- [391] “Kennedy Space Center Dataset,” <https://www.csr.utexas.edu/projects/rs/hrs/classify.html>, note = Accessed: 2023-01-30.
- [392] Z. Ge, G. Cao, X. Li, and P. Fu, “Hyperspectral image classification method based on 2d–3d cnn and multibranch feature fusion,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 13, pp. 5776–5788, 2020.
- [393] D. Hong, L. Gao, J. Yao, B. Zhang, A. Plaza, and J. Chanussot, “Graph convolutional networks for hyperspectral image classification,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 7, pp. 5966–5978, 2020.
- [394] C. Yu, R. Han, M. Song, C. Liu, and C.-I. Chang, “Feedback attention-based dense cnn for hyperspectral image classification,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–16, 2021.
- [395] L. Mou, X. Lu, X. Li, and X. X. Zhu, “Nonlocal graph convolutional networks for hyperspectral image classification,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 12, pp. 8246–8257, 2020.
- [396] Y. Chen, Y. Wang, Y. Gu, X. He, P. Ghamisi, and X. Jia, “Deep learning ensemble for hyperspectral image classification,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 12, no. 6, pp. 1882–1897, 2019.
- [397] L. Fang, W. Zhao, N. He, and J. Zhu, “Multiscale cnns ensemble based self-learning for hyperspectral image classification,” *IEEE Geoscience and Remote Sensing Letters*, vol. 17, no. 9, pp. 1593–1597, 2020.
- [398] H. Su, Y. Yu, Q. Du, and P. Du, “Ensemble learning for hyperspectral image classification using tangent collaborative representation,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 6, pp. 3778–3790, 2020.
- [399] J. Shi, T. Shao, X. Liu, X. Zhang, Z. Zhang, and Y. Lei, “Evolutionary multitask ensemble learning model for hyperspectral image classification,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 936–950, 2020.
- [400] 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 Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–15, 2022.
- [401] F. Ullah, Y. Long, I. Ullah, R. U. Khan, S. Khan, K. Khan, M. Khan, and G. Pau, “Deep hyperspectral shots: Deep snap smooth wavelet convolutional neural network shots ensemble for hyperspectral image classification,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 17, pp. 14–34, 2023.



computer vision.

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

Currently, he is working as a Post-Doctoral fellow with the College of Computer Science and Software Engineering at Shenzhen University, Shenzhen, Guangdong, China. His research interests include deep learning, hyperspectral image classification, and



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 from the Signals and Communication laboratory, University of Brescia, Italy, in 2016. He is currently working as an Assistant Professor in the Department of Computer Science, School of Engineering and Digital Sciences, Nazarbayev University, Kazakhstan.



Rehan Ullah Khan received the B.Sc. and M.Sc. degrees in information systems from the University of Engineering and Technology at Peshawar, in 2004 and 2006, respectively, and the Ph.D. degree from the Vienna University of Technology, Austria, in 2011. He is currently an Associate Professor with the IT Department, CoC, Qassim University, Saudi Arabia. His current research interests include machine learning, segmentation, and image-based object recognition.



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 in 2019, Ph.D. degree with the State Key Laboratory of Power Transmission Equipment and System Security and New Technology, School of Electrical Engineering, Chongqing University, Chongqing, China. In 2021, He has completed a Postdoctoral at the School of Engineering, Chulalongkorn University, Thailand.

He is currently working as an Associate Professor with the School of Computer Science, Chengdu University of Technology, China. He has published several papers in conferences and journals, such as IEEE Transactions on Power Electronics, IEEE Transactions on Magnetics, Applied Thermal Engineering, and IEEE Access.



Giovanni Pau (M'17) is an Associate Professor at the Faculty of Engineering and Architecture, Kore University of Enna, Italy. Prof. Pau received his Bachelor's degree in Telematic Engineering from the University of Catania, Italy, and both his Master's degree (cum Laude) in Telematic Engineering and Ph.D. from Kore University of Enna, Italy. Prof. Pau is the author/co-author of more than 80 refereed articles published in journals and conferences proceedings. He is a Member of the IEEE (Italy Section) and has been involved in the organization of

several international conferences as a session co-chair and technical program committee member. Prof. Pau serves/served as a leading Guest Editor in special issues of several international journals and is an editorial board member as Associate Editor of IEEE Access, Wireless Networks (Springer), EURASIP Journal on Wireless Communications and Networking (Springer), Wireless Communications and Mobile Computing (Hindawi), and Future Internet (MDPI). His research interests include Wireless Sensor Networks, Fuzzy Logic Controllers, Intelligent Transportation Systems, Internet of Things, and Network Security.



issues on the Internet, IoT, and IoV.

Salabat Khan received his M.Sc and MS degrees both in Computer Science from the University of Peshawar and International Islamic University, Islamabad, Pakistan, respectively. He completed his Ph.D. degree in Computer Science and Technology from the Beijing Institute of Technology, Beijing, China. Currently, he is working as a Post-Doctoral fellow in the College of Computer Science and Software Engineering at Shenzhen University, Shenzhen, Guangdong, China. His main research includes machine learning, deep learning, security and privacy