Review Article

Hyperspectral Image Classification: Potentials, Challenges, and Future Directions

Debaleena Datta [ID],[1] **Pradeep Kumar Mallick** [ID],[1] **Akash Kumar Bhoi** [ID],[2,3,4] **Muhammad Fazal Ijaz** [ID],[5] **Jana Shafi** [ID],[6] **and Jaeyoung Choi** [ID][7]

[1]*School of Computer Engineering, Kalinga Institute of Industrial Technology, Deemed to be University, Bhubaneswar 751024, India*
[2]*KIET Group of Institutions, Delhi-NCR, Ghaziabad-201206, India*
[3]*Directorate of Research, Sikkim Manipal University, Gangtok 737102, Sikkim, India*
[4]*AB-Tech eResearch (ABTeR), Sambalpur, Burla 768018, India*
[5]*Department of Intelligent Mechatronics Engineering, Sejong University, Seoul 05006, Republic of Korea*
[6]*Department of Computer Science, College of Arts and Science, Prince Sattam Bin Abdul Aziz University, Wadi Ad-Dawasir 11991, Saudi Arabia*
[7]*School of Computing, Gachon University, Seongnam-si 13120, Republic of Korea*

Correspondence should be addressed to Muhammad Fazal Ijaz; fazal@sejong.ac.kr and Jaeyoung Choi; jychoi19@gachon.ac.kr

Recent imaging science and technology discoveries have considered hyperspectral imagery and remote sensing. The current intelligent technologies, such as support vector machines, sparse representations, active learning, extreme learning machines, transfer learning, and deep learning, are typically based on the learning of the machines. These techniques enrich the processing of such three-dimensional, multiple bands, and high-resolution images with their precision and fidelity. This article presents an extensive survey depicting machine-dependent technologies' contributions and deep learning on landcover classification based on hyperspectral images. The objective of this study is three-fold. First, after reading a large pool of Web of Science (WoS), Scopus, SCI, and SCIE-indexed and SCIE-related articles, we provide a novel approach for review work that is entirely systematic and aids in the inspiration of finding research gaps and developing embedded questions. Second, we emphasize contemporary advances in machine learning (ML) methods for identifying hyperspectral images, with a brief, organized overview and a thorough assessment of the literature involved. Finally, we draw the conclusions to assist researchers in expanding their understanding of the relationship between machine learning and hyperspectral images for future research.

1. Introduction

Hyperspectral imagery is one of the most significant discoveries in remote sensing imaging sciences and technological advancements. Hyperspectral imagery (HSI) is the technology that depicts the perfect combination of Geographic Information System (GIS) and remote sensing. Besides, HSI has several advantages such as ecological protection, security, agriculture and horticulture applications, crop specification and monitoring, medical diagnosis, identification, and quantification [1]. RGB images are made up of three dimensions: width, height, and 3 color bands or channels consisting of color information, that is, red, green, and blue. They are stored as a 3D byte array that explicitly holds a color value for each pixel in the image; a combination of RGB intensities put down onto a color plane. However, in contrast, HSI comprises thousands of hypercubes and hence possesses a large resolution and an enormous amount of embedded information of all kinds—spectral, spatial, and temporal. This information enables various applications to

detect and characterize land covers, which are most significantly explored [2]. RGB images are captured by digital RGB cameras capable of characterizing objects only based on their shape and color. Moreover, the embedded information is minimal since only three visible bands are available in the human visibility range. The HSI, on the other hand, is captured by specialized airborne hyperspectral sensors placed on artificial satellites, that is, spectrometers. They have a broad range of scenes by acquiring large numbers of consecutive bands, not confined to the visible light spectrum and through a wider spectral band-pass. However, compared to the digital sensor that absorbs light in just three wide channels, a hyperspectral sensor's channel width is much narrower, making the spectral resolution and data volume much higher, resulting in hurdles to store, mine, and manage [3]. Furthermore, processing these data with a massive number of bands imposes many obstacles such as noise-causing image calibration, geometric distortion, noisy labels, and limited or unbalanced labeled training samples [4–6], that is, Hughes phenomenon and dimensionality reduction-related artifacts: overfitting, redundancy, spectral variability, loss of significant features between the channels, etc. [7].

Classifying HSIs is considered to be an intrinsically nonlinear problem [8], and the initial approach by linear-transformation-based statistical techniques such as principle component analytical methods, that is, principal component analysis (PCA) [9] and independent component analysis (ICA) [10]; the discriminant analytical methods, that is, linear [11] and fisher [12]; wavelet transforms [13]; and composite [14], probabilistic [15], and generalized [16] kernel methods, had shown promising outcomes. Still, their focus was limited to spatial information. They emphasized that the feature extractor techniques assisted by some basic random classifiers that lead to complexity in terms of cost, space, and time are not sufficiently accurate. After the success of these traditional methodical techniques assigned for HSI classification, researchers became keenly interested in applying the most recent emerging but not tedious computer-based methods that made the entire process smoother and vicinal to perfection. Study advancements suggest that the last decade can be considered the most escalating era regarding computer-based technologies due to the emergence of machine learning (ML). ML is an algorithmic and powerful tool that resembles the human brain's cognition. It simply represents a complex system by holding abstraction. Hence, it can reduce complexities and peep into the insights of the vast amount of HS data to fetch out the hidden discriminative features, both spectral and spatial [17]. Thus, it overcomes all the stumbling blocks to achieve the desired accuracy in identifying the classes that the objects of the target HSI data belong to. Hence, they act as all-in-one techniques that can serve the purpose without further assistance. Keeping this in mind, we conducted an extensive survey based on the various discriminative machine and deep learning (ML, DL) models for HSI. In most of the literature studies, the HSI datasets that are commonly used for landcover classification are AVIRIS Indian Pines (IP), Kennedy Space Center (KSC), Salinas Valley (SV), and ROSIS-03 University of Pavia (UP), along with less frequently used Pavia Center, Botswana, University of Houston (HU), etc. They are pre-refined and made publicly available on [18] for download and perform operations.

The motivation of our work is divided into three parts. First, a novel methodology is proposed for the review work that is entirely systematic and helps find the inspiration in forming the research gaps and embedded questions after going through a large pool of research articles. Second, this work focuses on the current advancements of ML technologies for classifying HSI, with their brief, methodical description and a detailed review of the literature involved with them. Finally, the inferences are drawn and help the researchers boost knowledge for their future research. The key contributions made to the research field on hyperspectral imagery by our novel effort are as follows:

(1) The thorough revision of the analytical and classification work carried out to date on HS imagery by employing ML/DL techniques.

(2) Emphasis on the categorized methods explored and practiced so far in an overly frequent manner. Also, it includes a brief interpretation of the most recent technologies and the highlighted hybrid techniques.

(3) An open knowledge base that acts as a reservoir of relevant information that is listed out that interprets all research on each mentioned technique in terms of their methodology, convenience and limitations, and future strategies. This illustration might administrate in making a proper choice of objective for further research on the field of HSIs.

(4) Explicit idea of the growth of interest in the concerned field that would attract researchers to invest themselves with a coherent, substantial specification (benefaction and drawbacks) of all the methods, individually, that contributes academically to the researchers about their favorable result and the difficulties for a chosen technique.

(5) A transitory rendition of the most recent research on HSIs signifies the currently adapted technologies as hot spots. Also, focus on the research areas about the interest that could apply to others, that is, the hybridized methods popular among researchers to address the problem and achieve the desired experimental results.

The rest of the article is arranged as follows: Section 2 briefly explains the constraints faced by the researchers in dealing with HSI; Section 3 represents the methodology for the research along with the motive behind this review; Section 4 describes seven ML techniques, namely, support vector machine (SVM), sparse representation (SR), Markov random field (MRF), extreme learning machine (ELM), active learning (AL), deep learning (DL), and transfer learning (TL); Section 5 shows up the complete summary of the literature review work in the form of answers to the research questions; Section 6 depicts the conclusions; and Section 7 explains the limitations and future work.

2. Constraints of HSI Classification

Since their emergence, several difficulties have caused issues in analyzing and performing operations on hyperspectral images. Initially, it suffered from spectroscopy technology due to the bad quality of hyperspectral sensors and poor quality with insufficient data. However, along with the advancement in applied science, things have come to ease, but there are still some well-known nondispersible hitches that need to be overcome. Some of them are stated as follows:

(a) Lack of high-resolution Earth observation (EO) noiseless images: During the initial stage of the discovery of spectrometers, they were not very efficient. Due to this, noises caused by water vapor, atmospheric pollutants, and other atmospheric perturbations modify the signals coming from the Earth's surface for Earth observations. Several efforts have been made over the last decades to produce high-quality hyperspectral data for Earth observation and develop a wide range of high-performance spectrometers that combines the power of digital imaging, spectroscopy, and extracting numerous embedded spatial-spectral features [19].

(b) Hindrances in the extraction of features: During data gathering, redundancy across contiguous spectral bands results in the availability of duplicated information, both spatially and spectrally, obstructing the optimal and discriminative retrieval of spatial-spectral characteristics [7].

(c) The large spatial variability and interclass similarity: The hyperspectral dataset collected contains unusable noisy bands due to mistakes in the acquisition that result in information loss in terms of the unique identity, that is, the spectral signatures and excessive intraclass variability. Furthermore, with poor resolution, each pixel comprises broad spatial regions on the Earth's surface, generating spectral signature mixing, contributing to the enhanced interclass similarity in border regions, thus creating inconsistencies and uncertainties for employed classification algorithms [19].

(d) Limitation of available training samples and insufficient labeled data: Aerial spectrometers cover significantly smaller areas, so they can only collect a limited number of hyperspectral data. That leads to the restriction of the number of training samples for classification models [20]. In addition, HSIs typically contain classes that correspond to a single scene, and available classification models' learning procedures require labeled data. However, labeling each pixel requires human skill, which is arduous and time-consuming [21].

(e) Lack of balance among interclass samples: The class imbalance problems, where each class sample has a wide range of occurrences, diminish the usefulness of many existing algorithms in terms of enhancing minority class accuracy without compromising majority class accuracy, which is a difficult task in and of itself [22].

(f) The higher dimensionality: Due to incorporating more information in multiple channels, such high-band pictures increase estimation errors. The curse of dimensionality is a significant drawback for supervised classification algorithms, as it significantly impacts their performance and accuracy [23].

The possible solutions to the above limitations that also represent the possible operations that are performed to analyze and comprehend the HSIs can be (1) technological advancement to make versatile and robust hardware for the spectrometers to capture the scenes more accurately, (2) spectral unmixing and resolution enhancement for better feature extraction and distinguishing capability of the embedded objects, (3) image compression-restoration and dimensionality reduction for addressing the high-dimensions and lack of data, and (4) use of robust classifiers that are capable of dealing with the above issues as well as promote fast computation ability [7].

These hurdles were very prominent for the methods that classify HSI based on the feature extrication from HSI. After ML/DL came into the scene, the operations on HSI became effortless as explicit feature extraction is not needed, and it has also many advantages such as great dealing with noise and time complexity. However, ML/DL acquires a few drawbacks in specific criteria [19], including parameter-tuning and numerous local minima problems in training procedures and compression [20] overfitting, optimization, and convergence problems despite many positive aspects.

3. Research Methodology

This section is divided into three categories that will assist in understanding the review procedure and its ambition.

3.1. Planning of the Review. Three systematic advances are utilized that comprise the planning behind our work. First, based on efficacy and frequency of applicability on classifying HSIs, seven most recently used ML techniques have been chosen in this article for review, which establishes the operational relationship and compatibility with the issue of categorizing the land covers of a particular scene captured as HSI. Second, this relationship provides all the shortfalls and benefits of those methods and their potential possibilities. Finally, we identified the limitations of our present review work and how to rectify them in the future.

3.2. Conducting the Review. The entire review work has been conducted in the following steps:

(a) Collection of literature: The literature studies have been collected based on the keywords: "Hyperspectral image classification," "Machine learning techniques," "Deep learning techniques," from the most relevant search engine, that is, Google (Google Scholar), which provides the scholarly articles for the concerned topic. These literature studies include
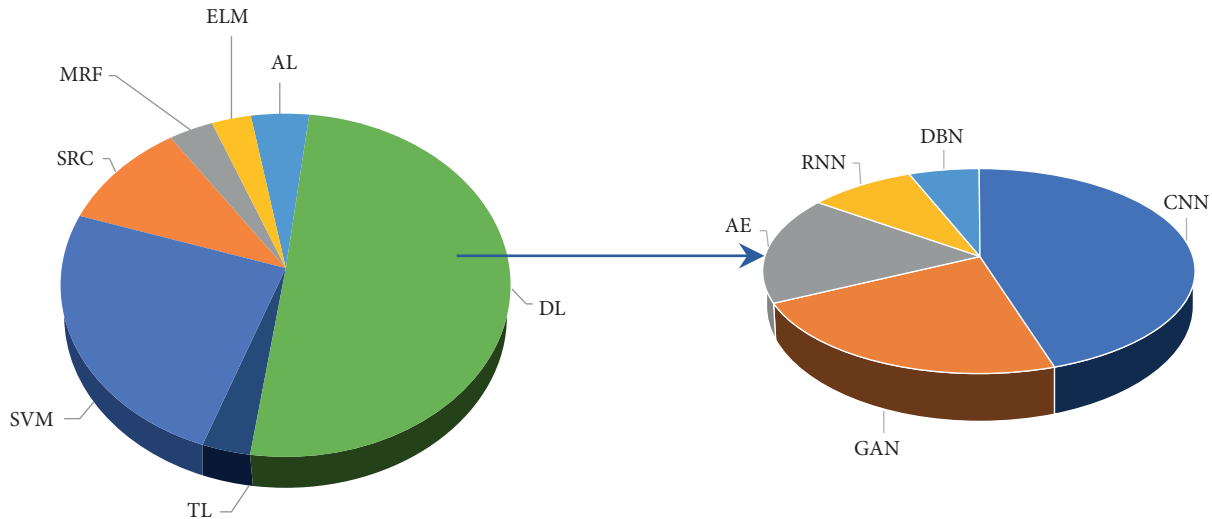
FIGURE 1: The statistical pie-charts of screened articles on ML/DL techniques used for HSI classification (source: SCI, SCIE, Scopus, WoS).

Web of Science (WoS), Scopus, SCI, and SCIE-indexed and SCIE-related articles, both journals and conferences. Several methods are utilized throughout the literature that assist the classification of hyperspectral data, out of which ML techniques seem to be more convenient and promising.

(b) Screening: The collected research papers depict raw data, sorted categorically according to the chronological order of the ML techniques used over the periods. The screening was accomplished based on the following constraints:

  (i) Time Period: The studies published in the range of 2010–2021 are included in this work. Studies published before 2010 are not included.
  (ii) Methodology: The studies on HSI's analytical operations (denoising, spectral unmixing, etc.) other than classifying the underlying land covers are rejected.
  (iii) *Type*: The studies that deal with the hyperspectral images of a particular land scene are considered, discarding the medical hyperspectral imagery, water reservoir, etc.
  (iv) Design of study: The studies comprising experimental outcomes and the elaboration of the models are accepted; other literary-based articles or review papers are only for primary knowledge gain.
  (v) The language used: The studies written in the English language are only considered.

Figure 1 represents the total number of the literary studies screened individually on each of the categories of chosen ML techniques in the form of pie-charts with a percent-wise pattern. Figure 2 is a standard graphical depiction of the number of most recent articles that we screened for each chosen ML-based method in the period ranging from 2015 to 2021.

(c) Selection: Out of all the papers screened based on the abovementioned criteria, a few most eligible are handpicked. The selection has been made keeping specific parameters: the modeling strategy and algorithm and its suitability with the modern technological scenario. The final result is the corresponding overall accuracy (COA) for each dataset used, preferably journals with a good citation index.

(d) Analysis and inference: These selected papers are thoroughly reviewed to determine their contribution, restrictions, and future propositions. Based on this analysis, the deductions are drawn to show the pathway of further research.

3.3. Research Investigations (RI). The analysis arises some of the queries:

RI 1: What is the significance of traditional ML and DL for analyzing HSI?

RI 2: How is ML/DL more impactful on HSI than other non-ML strategies?

RI 3: What are the advantages and challenges faced by the researchers for the chosen ML/DL-based algorithm for HSI classification?

RI 4: What are the emerging literary works of ML/DL on HSI classification in the year 2021?
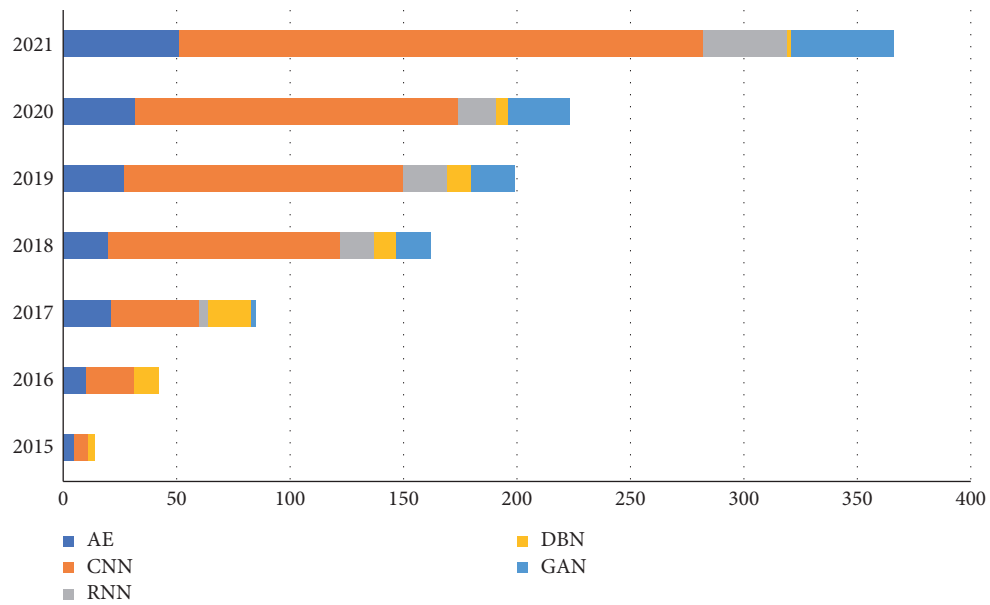
RI 5: How are ML- and DL-based hybrid techniques helping scientists in HSI classification?

RI 6: What are the latest emerging techniques associated with addressing classifying HSIs?

3.4. Datasets. The HSI datasets are pre-refined and made publicly available for download and perform operations. There are six datasets that are described here in a concise manner:
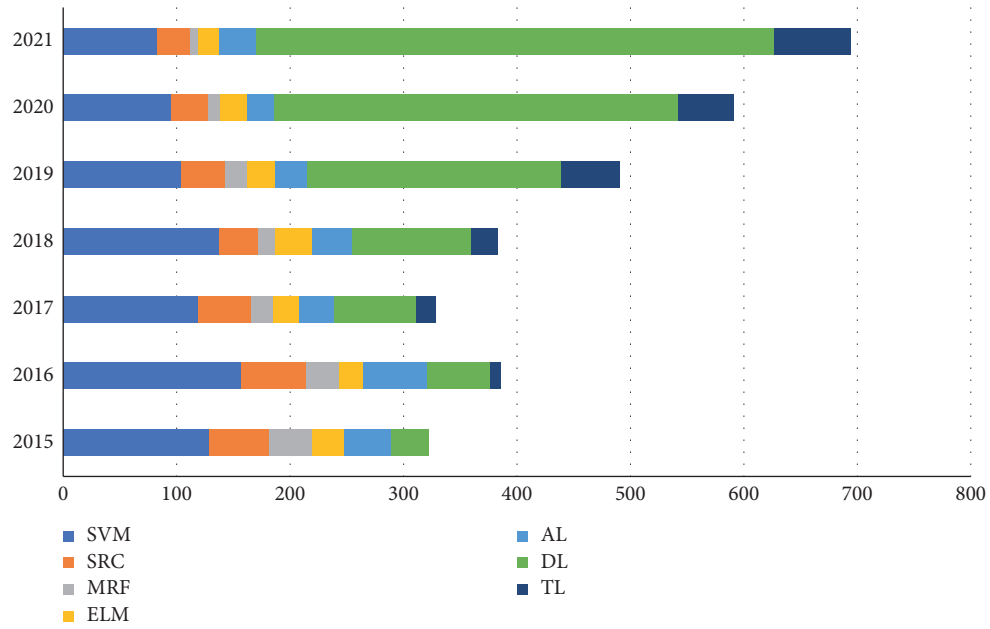
(a)

(b)

Figure 2: The statistical bar graph of screened articles on ML/DL techniques used for HSI classification from 2015 to 2021 (source: SCI, SCIE, Scopus, WoS): (a). ML. (b). DL.

(i) AVIRIS Indian Pines: This dataset was taken by airborne visible infrared imaging spectrometer (AVIRIS) sensor, on June 12, 1992. The scene captured here was Indian Pines test site in North-Western Indiana, USA, and contains an agricultural area exemplified by its crops of regular geometry and some irregular forest zones. It consists of 145 ∗ 145 pixels with a spectral resolution of 10 nm and a spatial resolution of 20 mpp and 224 spectral reflectance bands in the wavelength range

0.4–2.5 μm, out of which 24 noisy bans are removed due to low signal-to-noise ratio. The scene contains 16 different classes of land covers.

(ii) Salinas Valley: This scene was obtained by AVIRIS sensor over various agricultural fields of Salinas valley, California, USA, in 1998. The scene is characterized by a high spatial resolution of 3.7 mpp and a spectral resolution of 10 nm. The area is covered by 512 ∗ 217 spectral samples with a wavelength range of 0.4–2.5 μm. Out of 224 reflector
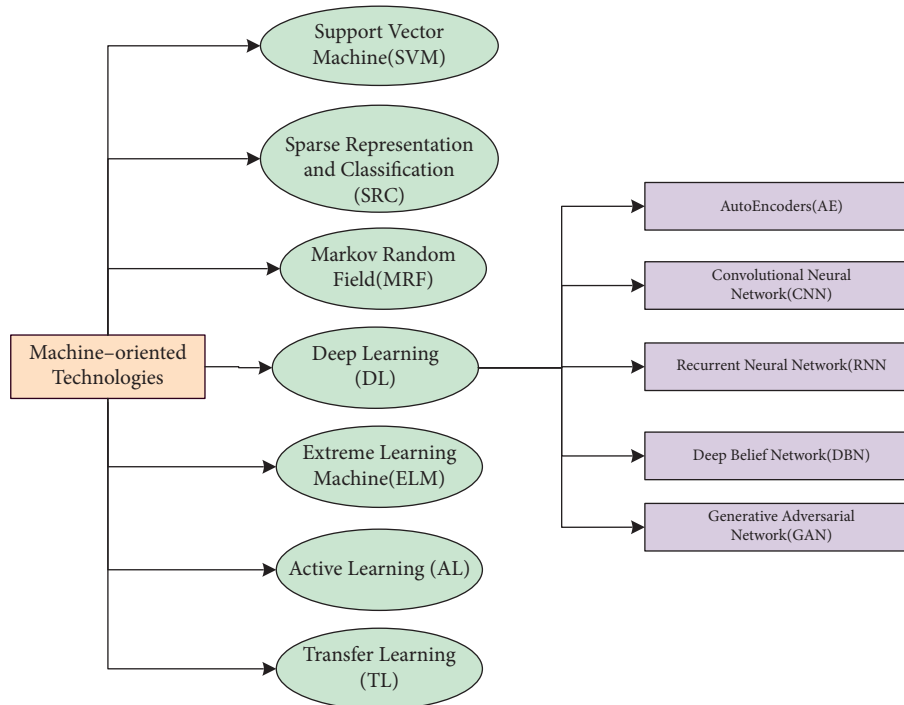
FIGURE 3: The categories of the eminent machine learning techniques used for HSI classification.

bands, 20 noisy bands are discarded due to water absorption coverage. The scene comprises 16 different land classes.

(iii) Pavia Center: This scene was captured by a reflective optics system imaging spectrometer (ROSIS-03) sensor during a flight campaign over Pavia, northern Italy. It possesses 115 spectral bands, out of which only 102 are useful. Its spectral coverage is 0.43–0.86 μm, with a spectral resolution of 4 nm and a spatial resolution of 1.3 mpp defined by 1096 ∗ 1096 pixels. There are 9 different land cover classes in the area.

(iv) Pavia University: This scene was also captured by the same sensor at the same time as Pavia center, over the University of Pavia in 2001. It has the same structural features as the Pavia center, only contrasting in considering 103 bands out of 115 bands with a size of 610 ∗ 340 are taken after discarding 12 noisy bands. The scene contains 9 classes with urban environmental constructions.

(v) Kennedy Space Center: This scene was acquired by NASA AVIRIS sensor over Kennedy Space Center, Florida, USA, on March 23, 1996. It was taken from an altitude of approximately 20 kilometres, having a spatial resolution of 18 kilometres and a spectral resolution of 10 nm. The wavelength range of the scene is 0.4–2.5 μm with the special size of 512 ∗ 614 pixels; 24 of 48 bands were removed for a low signal-to-noise ratio. The ground contains 13 predefined classes by the center personnel.

(vi) Botswana: The scene was obtained by the Hyperion sensor placed on the NASA EO-1 satellite over Okavango delta, Botswana, South Africa, on May 31, 2001. It has a special resolution of 30 metres and a spectral resolution of 10 nm while taken at an altitude of 7.7 kilometres. Out of 242 bands containing 1476 ∗ 256 pixels, with a wavelength range of 400–2500 nm, 97 bands are considered to be water-corrupted and noisy; hence, 145 remaining are useful. The scene comprises 14 land cover classes.

4. Machine Learning-Based Techniques for HSI Classification

ML technologies are not only intelligent and cognitive, but also their accuracy is skyrocketing due to their embedded mechanical abilities such as extraction, selection, and reduction of joint spatial-spectral features as well as contextual ones [24–26]. Moreover, the hidden dense layers with various allocated functions of the extensive networks work as intelligent learners by creating dictionaries or learning spaces to store deterministic information and then separate the landcover classes through its classification units [27–29]. The latest ML techniques that assist in classifying the hyperspectral data, that is, SVM, SRC, ELM, MRF, AL, DL, and TL, are shown categorically in Figure 3 and are discussed hereafter in detail.

4.1. Support Vector Machine (SVM). SVM is an innovative pattern-recognition technique rooted in the principle of statistical learning. The rudimentary concept of SVM-based training can unravel the ideal linear hyperplane so that the predicted classification error is mitigated, be it for binary or
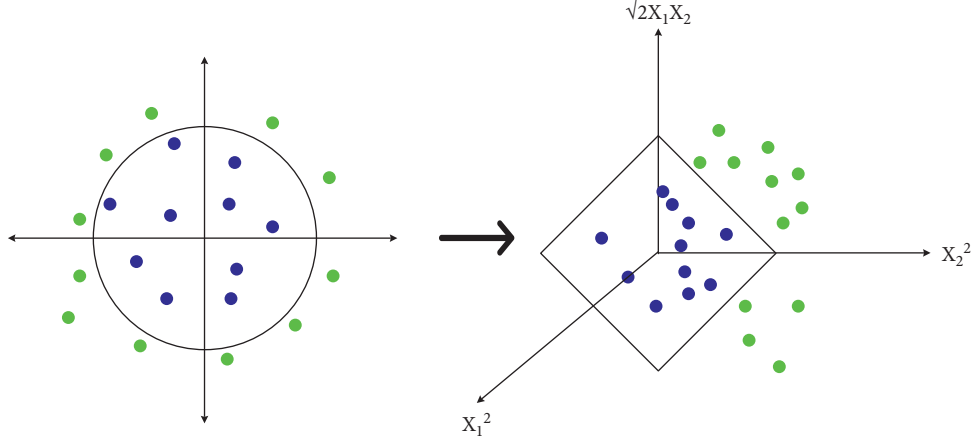
FIGURE 4: Classification strategy by multiclass SVM.

multiclass purposes [30], as depicted in Figure 4. For linearly separable binary classification, let (x_i, y_i) be the standard set of linearly separating samples with $x \in (R)^N$ and $y \in \{-1, +1\}$. The universal formula of linear decision function in n-dimensional space with the classification hyperplane is

$$g(x) = w^T.x + b = 0, \tag{1}$$

where w is the weight directional vector and b is the slope of the hyperplane. A separating hyperplane with margin $2/\|w\|$ in the canonical form must gratify the following constraints:

$$y_i\left[\left(w^T.x_i\right) + b\right] \geq 1. \tag{2}$$

For multiclass scenarios, we presumably transform the datapoints to S, a probable infinite-dimensional space, by a mapping function ψ defined as $\psi(x) = (x_1^2, x_2^2, \sqrt{2}x_1x_2)$, $\mathbf{x} = (x_1, x_2)$. Linear operations performed in S resemble nonlinear processes in the original input space. Let $K(x_i, x_j) = \psi(x_i)^T\psi(x_j)$ be the kernel function, which remaps the inner products of the training dataset.

Constructing SVM requires values of the constants, that is, Lagrange's multipliers, $\alpha = (\alpha_1, \ldots, \alpha_N)$ so that

$$P(\alpha) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{N} \alpha_i \alpha_j y_i y_j K(x_i.x_j). \tag{3}$$

is maximized with the constraints with respect to α:

$$\sum_{i=1}^{N} \alpha_i y_i = 0, \quad \alpha_i \geq 0 \text{ for all } \alpha_i. \tag{4}$$

Because most α_i are supposedly equal to zero, samples conforming to nonzero α_i are support vectors. Conferring to the support vectors, the modified optimally ideal classification function is

$$f(x) = \sum_{i=1}^{N} \alpha_i y_i K(x_i.x_j) + b. \tag{5}$$

The application of SVM for classifying HSI started two decades ago [31, 32]. Focusing on the potentially critical issue of applying binary SVMs [33], fuzzy-based SVM [34] as fuzzy input-fuzzy output support vector machine (F2-SVM), SVM evolved to dimensionality reduction and mixing of morphological details [35]. It also assisted particle swarm optimization (PSO) [36] and wavelet analysis with semi-parametric estimation [37], as the classifier "wavelet SVM" (WSVM). Table 1 summarizes the research carried out so far for the classification purpose of HSI using SVM.

4.2. Sparse Representation and Classification (SRC).
Sparse method depends on dictionary learning that enhances and rectifies the values of parameters based upon the current training observations while accumulating the knowledge of the previous observations prior. It then generates the sparse coefficient vector using sparse coding. This method is supremely efficient as it embeds dictionary learning to extract rich features embedded inside the HSI dataset. SR can classify images pixelwise by representing the patches around the pixel with a linear combination of several elements taken from the dictionary. The generalization of SRC called multiple SRC (mSRC) has three chief parameters—patch size, sparsity level, and dictionary size. Dictionary learning is the first step for sparse, using K-SVD algorithm. Let $Y = [y_1, y_2, \ldots, y_N]$ be a matrix of L2-normalized training samples $y_i \in R^m$ [45–47].

The size of patches around the pixel is

$$\min_{D,B} \|Y - DB\|_F^2 \text{ such that } \|b_i\| \quad 0 \leq S, \text{ for all } i, \tag{6}$$

where D is a member of $\mathbf{R}^{mXn}$ is the learned over a complete dictionary, with $n > m$ atoms, $B = [b_1, b_2, \ldots, b_m]$ represents the matrix of corresponding sparse coding vectors $b_i \in \mathbf{R}^n$, and $\|\cdot\|_F$ is the Frobenius norm. Sparsity S limits the number of nonzero coefficients in each b_i. The next step sparse coding is provided with dictionary D and represents y as a linear combination of $y = D\hat{b}$ where $\hat{b}$ is sparse. For the final classification step, suppose for each class $j \in \{1, \ldots, M\}$ of an image, a dictionary D_i is trained. Then, the classification of a new patch y_{test} is achieved by estimating a representation error. The class assignments rule [47] is calculated through a pseudoprobability measure $P(C_j)$ for each class error E_j as

TABLE 1: Summary of review of HSI classification using SVM.

Year	Method used	Dataset and COA	Research remarks and future scope
2011	Multiclass SVM [38]	San Diego3—98.86%	Outperforms traditional SVM and deals better with Hugh's effect
2012	Fuzzy decision tree-support vector machine (FDT-SVM) [39]	Washington DC mall—94.35%	Efficient testing accuracy truncated computational and storage demand, understandable edifices, and reduction of Hugh's effect
2014	Semi-supervised SVM kernel-spectral fuzzy C-means (KSFCM) [40]	IP—98.52%	Enhanced classification and clustering by fully exploring both labeled and unlabeled samples
2014	SVM-radial basis function (SVM-RBF) [41]	IP—88.7%, UP—94.7%	Outperforms other existing kernel-based methods
2015	Regional kernel-based SVM (RKSVM) [42]	UP—95.40%, IP—92.55%	Outperforms pixel-point-based SVM-CK
2017	Multiscale segmentation of super-pixels (MSP-SVMsub) [43]	UP: MSP-SVMsub—97.57%, IP: MSP-SVMsub-95.28%	Solving classic OBIC-based methods with difficulties determining the appropriate segmentation size reduces the Hughes phenomenon
2018	Extended morphological profiles (EMP), differential morphological profiles (DMP), Gabor filtering with SVM [44]	UP: MFSVM-GF—98.46%, IP: MFSVM-GF—98.01%	Outruns several advanced classifiers: SVM, super-pixel-based SVM, SVM-CK, multifeature SVM, EPF
2019	SVM-PCA [24]	IP—91.37%, UP—98.46%	Outperforms Naïve Bayes, decision tree k-NN

$$j* = \arg\max_j P(C_j), \quad \text{where,} \, P(C_j) = \frac{1}{M-1} \frac{\sum_{k=1, k \neq j}^{M} E_k}{\sum_{k=1}^{M} E_k}. \quad (7)$$

mSRC obtains residuals of disjoint sparse representation of y_{test} for all classes j. Each dictionary D_j is updated by eliminating nonzero atoms from $\widehat{b}_j$ after each of k iterations and y_{test} is assigned to the class, using Q total iterations:

$$D_j = \arg\max_j \sum_{k=1}^{Q} P_k(C_j). \quad (8)$$

Sparse representation is an essential and efficient machine-dependent method in many areas, including denoising, restoration, target identification, recognition, and monitoring. It may grow even more vital when associated with logistic regression, adaptivity, and super-pixels to extricate the joint features globally and locally. SR has a very high potential of being associated with methods such as PCA, ICA, Markov random fields, conditional random fields, extreme learning machines, and DL methods such as CNN and graphical convolutional network. Table 2 gives a summary of the research performed so far for the classification purpose of HSI employing SRC.

4.3. Markov Random Field (MRF).

MRF describes a set of random variables satisfying Markov probability, depicted by undirected graphs. It is similar to the Bayesian network but, unlike it, undirected and cyclic. An MRF is represented as a graphical model of a joint probability distribution defined in Figure 5. The undirected graph of MRF, $G = (V, E)$, in which V is the nodes representing random variables.

Based on the Markov properties [57], the neighborhood set N_c of a node c is defined as

$$N_c = \{c \in V \,|\, (c, d) \in E\}. \quad (9)$$

The conditional probability of Y_c decides the joint distribution of Y as

$$P(Y_c | Y_v - Y_c) = P(Y_c | Y_{\text{Nc}}). \quad (10)$$

To prosper the construction, the graph **G** absorbs a Gibbs distribution all over the maximum cliques (C) in **G**:

$$P(y) = \prod_{m \in C} \psi_m(y_m) = \frac{1}{Z} e^{-1/T \sum_{m \in C} V_m(y_m)}, \quad (11)$$

where Z is the partition function. Therefore, equation (11) can be rewritten as

$$P(y) = \frac{1}{Z} e^{-1/TU(y)}, \quad (12)$$

where T is the temperature, whose value is generally 1, and $U(y) = \sum_{m \in C} V_m(y_m)$ represents the energy.

Markov models depict the stochastic method that is represented by a graph made of circles has an acute advantage of not considering the past states for all upcoming future states for a random alterable dataset such as HSIs. The variants of Markov random fields are adaptive, hierarchical, cascaded, and probabilistic, a blend of Gaussian mixture model, joint sparse representation, transfer learning, etc., whose outcomes are pretty victorious. Hidden Markov random fields are highly suitable for the unsupervised classification of HSIs where the model parameters are estimated to make each pixel belong to its appropriate cluster [58], leading to the precise classification. Table 3 lists out the research carried out so far for the classification purpose of HSI employing MRF.

4.4. Extreme Learning Machine (ELM).

An efficacious learning algorithm based on single hidden layer feedforward neural network (SLFNN), it is applied to classify patterns and regression. Let $(x_i, p_i) \in \mathbf{R}^n \, X \, \mathbf{R}^m$ be N arbitrarily perceptible samples where $x_i = [x_{i1}, \ldots, x_{in}]^{\text{T}} \in \mathbf{R}^n$ and $p_i = [p_{i1}, \ldots, p_{im}]^{\text{T}} \in \mathbf{R}^m$ [72]. The standard SLFNN having $\widehat{N}$ hidden nodes and $f(x)$ as activation function is approached mathematically as

$$\sum_{i=1}^{\widehat{N}} \alpha_i f_i(x_i) = \sum_{i=1}^{\widehat{N}} \alpha_i f(w_i.x_j + b_i) = O_j; \quad j = 1, \ldots, N. \quad (13)$$

Table 2: Summary of review of HSI classification using sparse representation.

Year	Method used	Dataset and COA	Research remarks and future scope
2013	Kernel sparse representation classification (KSRC) [45]	IP—96.8%, UP—98.34%, KSC—98.95%	Lacks in devising automatic window size collection of spatial image quality, and filtering degree of class spatial relations
2014	Multiscale adaptive sparse representation (MASR) [46]	UP—98.47%, IP—98.43%, SV—97.33%	MASR outperformed the JSRM single-scale approach and several other classifiers on classification maps and accuracy The structural dictionary desired to be more inclusive and trained by discriminative learning algorithms
2015	Sparse multinomial logistic regression (SMLR) [47]	IP—97.71%, UP—98.69%	Being a pixelwise supervised method, its performance is better than other contemporary methods The model can be improved via more technical validations, exploitation of MRF, and structured sparsity-inducing norm that enhances the interpretability, stability, and identity of the model learned
2015	Super-pixel-based discriminative sparse model (SBDSM) [377]	IP—97.12%, SV—99.37%, UP—97.33%, Washington DC mall—96.84%	The advantages of this model lie in harnessing spatial contexts effectively through the super-pixel concept, which is better in performance speed and classification accuracy Determination of a supplementary and systematic way to adjust the count of super-pixels to various conditions and apply SR to other remote sensing practices
2015	Shape-adaptive joint sparse representation classification (SAJSRC) [48]	IP—98.45%, UP—98.16%, SV—98.53%	Local area shape-adapted for every test pixel rather than a fixed square window for adaptive exploration of spatial PCs, making the method outperforms other corresponding methods Region searching based on shape-adaption can be used instead of the reduced dimensional map to reconnoiter complete spatial information of the actual HSI
2017	Multiple-feature-based adaptive sparse representation (MFASR) [49]	IP—97.99%, UP—98.39%, Washington DC mall—97.26%	SA regions' full utilization of all embedded joint features makes the method superior to some cutting-edge approaches Enhancement of the proposed method in the future by selecting features automatically and improving dictionary learning to reduce the computational cost
2018	Weighted joint nearest neighbor and joint sparse representation (WJNN-JSR) [50]	UP—97.42%, IP— 93.95%, SV—95.61%, Pavia center—99.27%	The model was improved using the Gaussian weighted method and incorporates the conventional test pixel area to achieve a new measure of classification knowledge: The Euclidean-weighted joint size Creating more effective approaches to applying the system and further increasing classification accuracy are taken as future work
2019	Log-Euclidean kernel-based joint sparse representation (LogEKJSR) [51]	IP—97.25%, UP—99.06%, SV—99.36%	Specializes in extracting covariance traits from a spatial square neighborhood to calculate the analogy of matrices with covariances employing the conventional Gaussian form of Kernel Creation of adaptive local regions using super-pixel segmentation methods and learning the required kernel using multiple kernel learning methods

TABLE 2: Continued.

Year	Method used	Dataset and COA	Research remarks and future scope
2019	Multiscale super-pixels and guided filter (MSS-GF) [52]	IP—97.58%, UP—99.17%	Effective spatial and edge details in his, various regional scales to build MSSs to acquire accurate spatial information, and GF improved the classification maps for near-edge misclassifications Additional applications of efficient methods to extract local features and segment super-pixels are added as future work
2019	Joint sparse representation—self-paced learning (JSR-SPL) [53]	IP—96.60%, SV—98.98%	The findings are more precise and reliable than other JSR methods
2019	Maximum-likelihood estimation based JSR (MLEJSR) [54]	IP—96.69%, SV—98.91%, KSC—97.13%	The model is reliable in terms of outliers
2020	Global spatial and local spectral similarity-based manifold learning-group sparse representation-based classifier (GSLS-ML-GSRC) [55]	UP—93.42%, Washington DC mall—91.64%, SV—93.79%	The said fusion makes the method outperform other contemporary methods focused on nonlocal or local similarities
2020	Sparse-adaptive hypergraph discriminant analysis (SAHDA) [56]	Washington DC mall—95.28%	Effectively depict the multiple complicated aspects of the HSI and will be considered for future spatial knowledge

Here, $w_i = [w_{i1}, \ldots, w_{in}]^T$ gives the weight vector establishing the connection between input nodes and i^{th} is the hidden node and $\alpha_i = [\alpha_{i1}, \ldots, \alpha_{im}]^T$ represents the weight vector connecting between output node O_j with the ith hidden node, and $w_i.x_j$ represents the inner product. The zero error for N samples can be written in the matrix form as

$A\alpha = P$, where A $(w_1, \ldots, w_{\widehat{N}}, b_1, \ldots, b_{\widehat{N}}, x_1, \ldots, x_N)$ is the neural network hidden layer output matrix, and the ith is hidden node output with respect to $x_1, \ldots, x_N$; the ith column of A represents x_N inputs. The training of SLFNN is based on finding specific α, w_i, and $b_{i,}$ $(i = 1, \ldots, \widehat{N})$ [73] such that

$$\left\| A\left(w_1, \ldots, w_{\widehat{N}}, b_1, \ldots, b_{\widehat{N}}, x_1, \ldots, x_N\right)\alpha - P \right\| = \min_{w,\alpha,b} \left\| A\left(w_1, \ldots, w_{\widehat{N}}, b_1, \ldots, b_{\widehat{N}}, x_1, \ldots, x_N\right)\alpha - P \right\|. \tag{14}$$

This equation denotes the cost function with a depreciation. By using gradient-based algorithms, the set of weights (α_i, w_i) and biases b_i are attuned with epochs as

$$w_k = w_{k-1} - \eta \frac{\delta U(W)}{\delta W};$$

$$U = \sum_{k=1}^{N} \left(\sum_{j=1}^{\widehat{N}} \alpha_j f\left(w_j. x_k + b_j\right) - P_k \right)^2. \tag{15}$$

The learning rate η must be accurate for better convergence and $\widehat{N} << N$ for better generalization performance.

Extreme learning methods proposed overcoming the disadvantage of a single hidden layer feedforward neural network and improving learning ability and generalization performance. It is a supervised method but is highly recommended to get an extension to its semi-supervised and unsupervised versions for dealing with the huge amount of data such as HSIs, which are primarily unlabeled and suffering from lack of training samples. Great potential lies with its other variants than those mentioned here, [74] of ELM, like two-hidden layer ELM, multilayer ELM, feature mapping-based ELM, incremental ELM, and deep ELM to become superior and achieve victorious precision in classifying

HSIs. Table 4 underneath provides the summary of the research executed so far for the classification purpose of HSI utilizing ELM.

4.5. Active Learning (AL). It is a special type of the supervised ML approach to build a high-performance classifier while minimizing the size of the training dataset by actively selecting valuable data points. The general structure of AL can be understood from Figure 6. There are three categories of AL—stream-based selective sampling, that is, where each unlabeled dataset is enquired for a certain label whether to assign a query or not; pool-based sampling; that is, the whole dataset is under consideration before selecting the best set of queries; and membership query synthesis; that is, it involves data augmentation to create user selected labeling. The decision to select the most informative data points depends on the uncertainty measure used in the selection. In an active learning scenario, the most informative data points are those the classifier is least sure about. The uncertainty measures for datapoints x [88] are

Least Confidence (LC): responsible for selecting the classifier's data point is least certain about the chosen class. With y^* as the most likely label sequence and φ as the learning model, LC is represented as
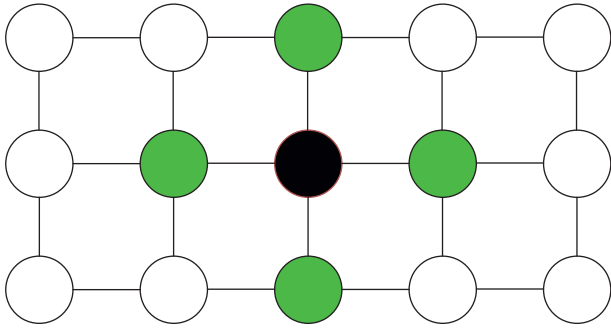
FIGURE 5: Given the green nodes, the black node is independent of other nodes.

$$S_{\text{LC}}(x) = 1 - P(y * | x, \phi). \tag{16}$$

Smallest Margin Uncertainty (SMU): Represents the difference between classification probability of the most likely class (y_1*) and that of the second-best class (y_2*), written mathematically as:

$$S_{\text{SMU}}(x) = P_{\phi}(y_1 * | x) - P_{\phi}(y_2 * | x). \tag{17}$$

Largest Margin Uncertainty (LMU): Represents the difference between classification probability of most likely class (y_1*) and that of the least likely class ($y_{\min}$), written mathematically as:

$$S_{\text{LMU}}(x) = P_{\phi}(y_1 * | x) - P_{\phi}(y_{\min} * | x). \tag{18}$$

Sequence Entropy (SE): Detects the measure of disorder in a system; higher the entropy implies a more disordered condition. The denotation of SE is

$$S_{\text{SE}}(x) = -\sum_{\widehat{y}} P(\widehat{y}|x; \phi)\log P(\widehat{y}|x; \phi), \tag{19}$$

with $\widehat{y}$ ranging over all possible label sequences for input x.

Although not considered customary and coherent, AL is pretty much capable of reducing human effort, time, and processing cost for a large batch of unlabeled data. This method relies on prioritizing data that needs to be labeled in a huge pool of unlabeled data to have the highest impact on training. A desired supervised model keeps on being trained through active queries and improvising itself to predict the class for each remaining data point. AL is advantageous for its dynamic and incremental approach to training the model so that it learns the most suitable label for each data cluster [89]. Table 5 lists out the research performed so far for the classification purpose of HSI using AL.

4.6. Deep Learning (DL). Deep learning is the most renowned ML technology in application and accuracy terms. Although it is considered the next tread of ML, it also lends concepts from artificial intelligence. DL is the mother of algorithms that resemble human brain simulations, that is, creativity, enhanced analysis, and proper decision-making, based on pure or hybrid large networks for any given real-life

problem. It has enhanced the throughput of computer-based, especially unsupervised snags for the practical technology-based applications such as automated translation of machines, image reconstructions and classifications, computer vision, and automated analysis. [104] The basic structure of any DL model possesses a three-type-layered architecture: it contains one input layer through which input data are fed to the next layer(s) known as the intermediate hidden layer responsible for all the computations based on the problem given, which passes its generated data to the final layer, that is, the output layer, which provides the desired ultimate output. The steps involved in DL models are as follows: having proper knowledge and understanding of the problem, collecting the input database, selecting the most appropriate algorithm, training the model with the sample source database, and finally testing the target database [105].

DL models are more efficient and advantageous over other ML models due to the following reasons [19]:

(1) The capability to extract hidden and complicated structures from raw data is inextricably linked to their ability to represent the internal representation and generalize any form of knowledge.

(2) They have a wide range of data types that they can accommodate, for example, 2D imagery data and complex 3D data such as medical imagery and remote sensing. In addition, they can use HSI data's spectral and spatial domains in both standalone and linked ways [106–108].

(3) They provide architects a lot of versatility in terms of layer types, blocks, units, and depth.

(4) Furthermore, its learning approach can be tailored to various learning strategies, from unsupervised to supervised, with intermediate strategy.

(5) Additionally, developments in processing techniques, including batch partitioning and high-performance computation, especially on distributed and parallel architecture, have enabled DL models to find better opportunities and solutions when coping with enormous volumes of data [109].

The models that are broadly used for HSI classification are described as follows.

(a) Autoencoder (AE): AEs are the fundamental unsupervised deep model based on the backpropagation rule. AEs consist of two fragments: encoder, connecting the input vector to the hidden layer by a weight matrix; decoder, formed by the hidden layer output via a reconstruction vector tied by a specific weight matrix. SAEs are AEs with multiple hidden layers where the production of every hidden layer is fed to the successive hidden layer as input. It comprises three steps: (1) first AE trained to fetch the learned feature vector; (2) the former layer's feature vector is taken as input to the next layer, and this process is redone till the completion of training;

TABLE 3: Summary of review of HSI classification using MRF.

Year	Method used	Dataset and COA	Research remarks and future scope
2011	Adaptive-MRF (a-MRF) [59]	IP—92.55%	Handles homogeneous problem of "salt and pepper" areas and the possibility of overcorrection impact on class boundaries
2014	Hidden MRF and SVM (HMRF-SVM) [60]	IP—90.50%, SV—97.24%	Outperforms SVM and improves overall accuracy outcomes by nearly 8% and 3.2%, respectively
2014	Probabilistic SR with MRF-based multiple linear logistic (PSR-MLL) [61]	IP—97.8%, UP—99.1%, Pavia center—99.4%	Exceeds other modern contemporary methods in terms of accuracy
2014	MRF with Gaussian mixture model (GMM-MRF) [62]	UP(LFDA-GMM-MRF)-90.88% UP(LPNMF-GMM-MRF)—94.96%	Advantageous for a vast range of operating conditions and spatial-spectral information to preserve multimodal statistics GMM classificatory distributions are to be considered in the future
2011	MRF with sparse multinomial logistic regression classifier—spatially adaptive total variation regularization (MRF-SMLR-SpATV) [63]	UP—90.01%, IP—97.85%, Pavia center—99.23%	Efficient time complexity of the model Improvisation of the model by implementing GPU and learning dictionaries are the future agendas
2016	Multitask joint sparse representation (MJSR) and a stepwise Markov random filed framework (MSMRF) [64]	IP—92.11%, UP—92.52%	The gradual optimization explores the spatial correlation, which significantly improves the effectivity and accuracy of the classification
2016	MRF with hierarchical statistical region merging (HSRM) [65]	SVMMRF-HSRM: IP—93.10%, SV—99.15%, UP— 86.52%; MLRsubMRF-HSRM-IP—82.60%, SV—88.16%, UP—95.52%	Better solution to the technique of majority voting that suffers from the problem of scale choice Considering the spatial features in the spatial prior model of objects of the different groups in the future
2018	Integration of optimum dictionary learning with extended hidden Markov random field (ODL-EMHRF) [66]	ODL-EMHRF-ML-IP—98.56%, UP—99.63%; ODL-EMHRF-EM-IP—98.47%, UP—99.58%	The method has been proven to be better than SVM-associated EMRF
2018	Label-dependent spectral mixture model (LSMM) fused with MRF (LSMM-MRF) [67]	The Konka image—94.19%, the shipping scene—66.45%	Efficient unsupervised classification strategy that considers spectral information in mixed pixels and the impact of spatial correlation Enhanced theoretical derivations of EM steps
2019	Adaptive interclass-pair penalty and spectral similarity information (aICP2-SSI) along with MRF and SVM [68]	UP—98.10%, SV—96.40%, IP— 96.14%	Outperforms other MRF-based methods More efficient edge-preserving strategies, more spectral similitude, and class separable calculation methods as future research
2019	Cascaded version of MRF (CMRF) [69]	IP—98.56%, Botswana—99.32%, KSC—99.24%	Backpropagation tunes the model parameters and least computation expenses
2020	Fusion of transfer learning and MRF (TL-MRF) [70]	IP—93.89%, UP—91.79%	TL is taken to be very effective for HSI classification Future research for reducing the number of calculations involved in the existing
2020	MRF with capsule net (caps-MRF) [71]	IP—98.52%, SV—99.74%, Pavia center—99.84%	Ensures that relevant information is preserved, and the spatial constraint of the MRF helps achieve more precise model convergence The combination of CapsNet with several postclassification techniques

TABLE 4: Summary of review of HSI classification using ELM.

Year	Method used	Dataset and COA	Research remarks and future scope
2014	Ensemble extreme learning machines (E^2LM)-bagging-based ELMs (BagELMs) and AdaBoost-based (BoostELMs) [72]	UP—94.3%, KSC—97.71%, SV—97.19%	BoostELM performs better than kernel and other EL methods. Performance of other differential or nondifferentiable activation functions
2015	Kernel-based ELM—composite kernel (KELM-CK) [75]	IP—95.9%, UP—93.5%, SV—96.4%	Outperforms other SVM-CK-based models
2015	ELM's two-level fusions: feature-level fusion (FF-ELM) and mixing ELM classifier two levels of fusions: feature-level fusion (FF-ELM) [76]	FF-ELM: UP—98.11%, IP—92.93%, SV—99.12%; DF-ELM—UP—99.25%, IP—93.58%, SV—99.63%	Outperforms basic ELM models
2016	Hierarchical local-receptive-field-based ELM (HL-ELM) [77]	IP—98.36%, UP—98.59%	Surpasses other ELM methods in terms of accuracy and training speed
2017	Genetic-firefly algorithm with ELM (3FA-ELM) [78]	HyDice DC mall—97.36%, HyMap—95.58%	Low complexity (ELM), better adaptability, and searching capability (FA). Execution time needs to be reduced in future
2017	Local receptive fields-based kernel ELM (LRF-KELM) [79]	IP—98.29%	Outperforms other ELM models
2017	Distributed KELM based on MapReduce framework with Gabor filtering (DK-Gabor-ELMM) [80]	IP—92.8%, UP—98.8%	Outperforms other ELM models
2017	Loopy belief propagation with ELM (ELM-LBP) [81]	IP—97.29%	Efficient time complexity
2018	Mean filtering with RBF-based KELM (MF-KELM) [82]	IP—98.52%	The model offers the most negligible computational hazard
2018	Augmented sparse multinomial logistic ELM (ASMLELM) [83]	IP—98.85%, UP—99.71%, SV—98.92%	Improved classification accuracy by extended multi-attribute profiles and more SR
2018	ELM with enhanced composite feature (ELM-ECF) [84]	IP—98.8%, UP—99.7%, SV—99.5%	Low complexity and multiscale spatial feature for better accuracy. Incorporate feature-fusion technology
2019	Local block multilayer sparse ELM (LBMSELM) [85]	IP—89.31%, UP—89.47%, SV—90.03%	Performs anomaly and target detection. Reduced computational overhead and increased classification accuracy by inverse free; saliency detection and gravitational search
2019	ELM-based heterogeneous domain adaptation (EHDA) [25]	HU-DC —97.51%, UP-DC —96.63%, UP-HU —97.53%	Outperforms other HDA methods. Invariant feature selection
2019	Spectral-spatial domain-specific convolutional deep ELM (S^2CDELM) [86]	IP—97.42%, UP—99.72%	Easy construction with high training-testing speed. Merge of DL with ELM
2020	Cumulative variation weights and comprehensive evaluated ELM (CVW-CEELM) [87]	IP—98.5%, UP—99.4%	Accuracy achieved due to the weight determination of multiple weak classifiers. Multiscale neighborhood choice and optimized feature selection

(3) backpropagation is used after all the hidden layers have been trained to reduce the cost function and to update the weights is done with a named training set to obtain fine-tuning [110]. The architecture of SAE is depicted in Figure 7.

Let $x_n \in \mathbf{R}^m$; $n = 1, 2, \ldots, N$ represent the unlabeled input dataset, E_n be the hidden encoder vector computed by x_n, and y_n be the decoder vector of the output layer [111].

$$\text{Encoder}: E_n = g(W_i x_n + b_i); \quad (20)$$

g-> encoding function, W_i-> encoder weight matrix, b_i-> encoder bias vector.

$$\text{Decoder}: y_n = f(W_j E_n + b_j); \quad (21)$$

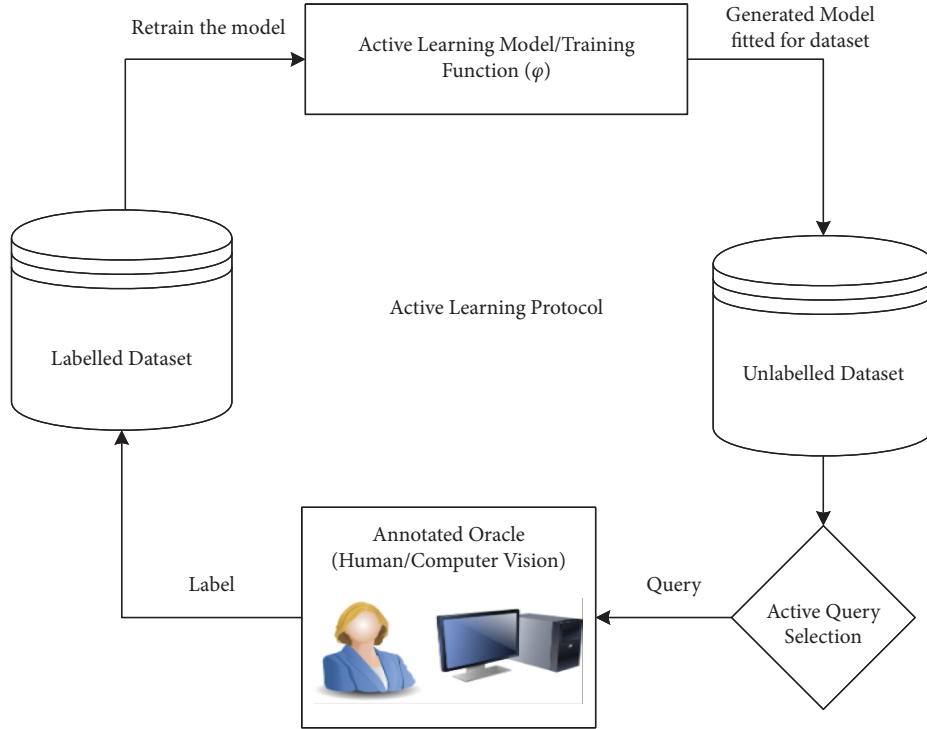f-> decoding function, W_j-> decoder weight matrix, b_j-> decoder bias vector.

FIGURE 6: Principle of active learning.

The reconstruction error in SAE is denoted as

$$\Phi(\Theta) = \arg\min_{\theta,\theta'} \frac{1}{N} \sum_{k=1}^{N} L(x^k, y^k), \text{ where the Loss function is } L(x^k, y^k) = \|x - y\|^2. \tag{22}$$

AEs are unsupervised neural networks that embed several convolutional hidden layers based on nonlinear activation functions and transformations [112]. There are high risks of data loss during training, but it handles the model well for specific data types through specialized training. There are AEs for every purpose such as convolutional, sparse, variational, deep, contractive, and denoising applied for data compression, noise removal, feature extraction, image augmenting, and image coloring. AE inevitably provides a vast platform for further research on its various applicability and its capability to participate in hybridization. Table 6 describes a few research works in the aspect of AEs.

(b) *Convolutional Neural Network (CNN)*: It is a famous deep neural network that works like a human visual cortex with many interconnected layers applied widely in image, speech, and signal processing. It assigns learnable and modifiable weights and biases to the input image to identify various objects or patterns with differentiable features. As shown in

Figure 8, each layer of CNN possesses filtering capabilities with ascending complexities: the first layer learns filtering corners and edges; intermediate layers learn object parts filtering; and the last layer learns filtering out the entire object in different locations and shapes. The comparison between the layers in terms of several parameters is shown in Table 7. It consists of four layers [117, 118]:

(1) *Convolution*: This operation is the cause of the naming of CNN, that is, a dot product of the original pixel values with weights identified in the filter or kernel of the image. The findings are compiled into one number representing all the pixels found in the filter. Assuming $\mathbf{I}$ be the hyper-input-cube of dimension $p \times q \times r$ where $p \times q$ denotes the spatial size of $\mathbf{I}$ with r number of bands, and i_k is the kth feature map of $\mathbf{I}$. Let d number of filters be present in each convolutional layer, and weight W_m and bias b_m represent the mth filter. The mth convolutional layer output with transformation function g is denoted as

TABLE 5: Summary of review of HSI classification using active learning.

Year	Method used	Dataset and COA	Research remarks and future scope
2008	AL with expectation-maximization-binary hierarchical classifier (BHC-EM-AL) and maximum-likelihood (ML-EM-AL) [90]	Range: KSC-90-96%, Botswana—94-98%	Better learning levels than the random choice of data points and an entropy-based AL Measurement of the efficacy of the active learning-based knowledge transfer approach while systematically increasing the spatial/temporal segregation of the data sources
2010	Semi-supervised-segmentation with AL and multinomial logistic regression (MLR-AL) [91]	IP—79.90%, SV—97.47%	Innovative mechanisms for selecting unlabeled training samples automatically, AL to enhance segmentation results Testing the segmentation in various scenarios influenced by limited a priori accessibility of training images
2013	Maximizer of the posterior marginal by loopy belief propagation with AL (MPM-LBP-AL) [92]	IP—94.76%, UP—85.78%	Improved accuracy than previous AL applications Use parallel-computer-architectures such as commodity—clusters or GPUs to build computationally proficient implementation
2015	Hybrid AL-MRF, that is, uncertainly sampling breaking ties (MRF-AL-BT), passive selection approach random sampling (MRF-AL-RS), and the combination (MRF-AL-BT + RS) [93]	IP—94.76%, UP—85.78% (MRF-AL-RS provides the highest accuracies)	Outperforms conventional AL and SVM AL methods due to MRF regularization and pixelwise output Merge the model with other effective AL methods and test them with a limited number of training samples
2015	Integration of AL and Gaussian process classifier (GP-AL) [94]	IP—89.49%, Pavia center—98.22%	Empirical autonomation of AL achieves reasonable accuracy Adding diversity criterion to the heuristics and contextual information with the model and reducing computation time
2016	AL with hierarchical segmentation (HSeg) tree: adding features and adding samples (Adseg_AddFeat + AddSamp) [95]	IP—82.77%, UP—92.23%	Outruns several baseline methods-selecting appropriate training data from already existing labeled datasets and potentially decreasing manual laboratory labeling Reduce the computational time that limits its applicability on large-scale datasets
2016	Multiview 3D redundant discrete wavelet transform-based AL (3D-RDWT-MV-AL) [96]	HU—99%, KSC—99.8%, UP—95%, IP—90%	The precious method as a combination of an initial process with AL, improved classification
2017	Discovering representativeness and discriminativeness by semi-supervised active learning (DRDbSSAL) [97]	Botswana—97.03%, KSC—93.47%, UP—93.03%, IP—88.03%	Novel approach with efficient accuracy
2017	Multicriteria AL [98]	KSC—99.71%, UP—99.66%, IP—99.44%	Surpasses other existing AL methods regarding stability, accuracy, robustness, and computational hazard A multi-objective optimization strategy and the usage of advanced attribute-based profile features
2018	Feature-driven AL associated with morphological profiles and Gabor filter [99]	IP—99.5%, UP—99.84%, KSC—99.53% (Gabor-BT)	A discriminative feature space is designed to gather helpful information into restricted samples
2018	Multiview intensity-based AL (MVAL)-multiview intensity-based query-representative strategy (MVIQ-R) [100]	UP—98%, Botswana—99.5%, KSC—99.9%, IP—95%	Focus on pixel intensity obtains unique feature and hence better performance Selection of combination of optimal attribute features
2019	Super-pixel with density peak augmentation (DPA)-based semi-supervised AL (SDP-SSAL) [101]	IP—90.08%, UP—85.61%	Novel approach proposed based on super-pixels density metric Development of a pixelwise solution to produce super-pixel-based neighborhoods

Table 5: Continued.

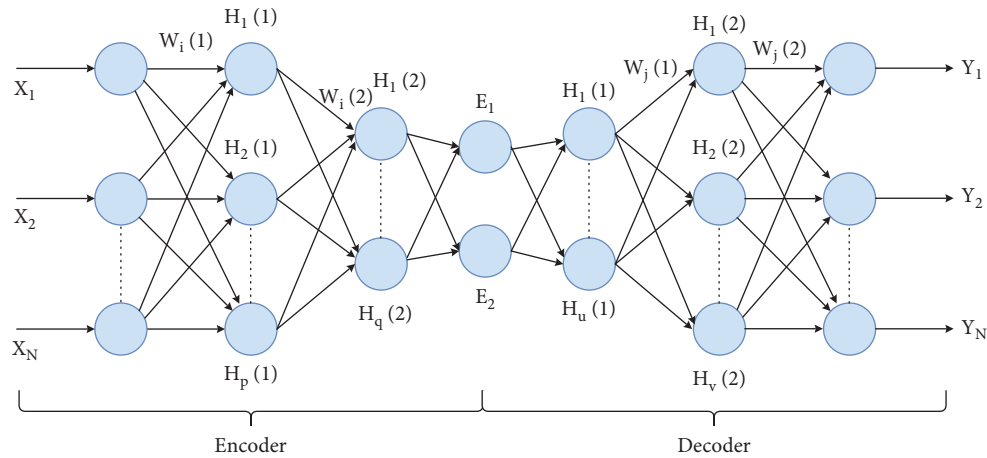| Year	Method used	Dataset and COA	Research remarks and future scope
2020	Adaptive multiview ensemble spectral classifier and hierarchical segmentation (Ad-MVEnC_Spec + Hseg) [102]	KSC—97.63%, IP—87.1%, HU—93.3%	Enhancement in the view sufficiency, and promotion of the disagreement level by the dynamic view, provides lower computational complexity due to parallel computing
2020	Spectral-spatial feature fusion using spatial coordinates-based AL (SSFFSC-AL) [103]	IP—100%, UP—98.43%	High running speed can successfully address the "salt and pepper" phenomenon but drops a few if similar class samples are distributed in different regions differently The sampling weight parameter conversion to an adaptive parameter is adjusted adaptively as the training samples are modified

Figure 7: The network structure of stacked autoencoders; input X-to-E is the encoding phase; E-to-output Y is the decoding phase.

Table 6: Summary of the review of HSI classification using deep learning—AE.

| Year	Method used	Dataset and COA	Research remarks and future scope
2013	Autoencoders (AE) [110]	Error rate: KSC—4%, Pavia city—14.36%	This article opened a considerable doorway of research, including other deep models for better accuracy
2014	Stacked autoencoder and logistic regression (SAE-LR) [113]	KSC—98.76%, Pavia city—98.52%	Highly accurate in comparison to RBF-SVM and performs testing in optimized time limit than SVM or KNN but fails in training time efficiency
2016	Spatial updated deep AE with collaborative representation-based classifier (SDAE-CR) [114]	IP—99.22%, Pavia center—99.9%, Botswana—99.88%	Highly structured in extracting high specialty deep features and not the hand-crafted ones and accurate Improving the deep network architecture and selection of parameters
2019	Compact and discriminative stacked autoencoder (CDSAE) [115]	UP—97.59%, IP—95.81%, SV—96.07%	Efficient in dealing with feature space in low dimension, but the computation cost is high as per architecture size
2021	Stacked autoencoder with distance-based spatial-spectral vector [116]	SV—97.93%, UP—99.34%, surrey—94.31%	Augmentation of EMAP features with the geometrically allocated spatial-spectral feature vectors achieves excellent results. Better tuning of hyperparameter and more powerful computational tool required Improving the training model to become unified and classified in a more generalized and accurate way
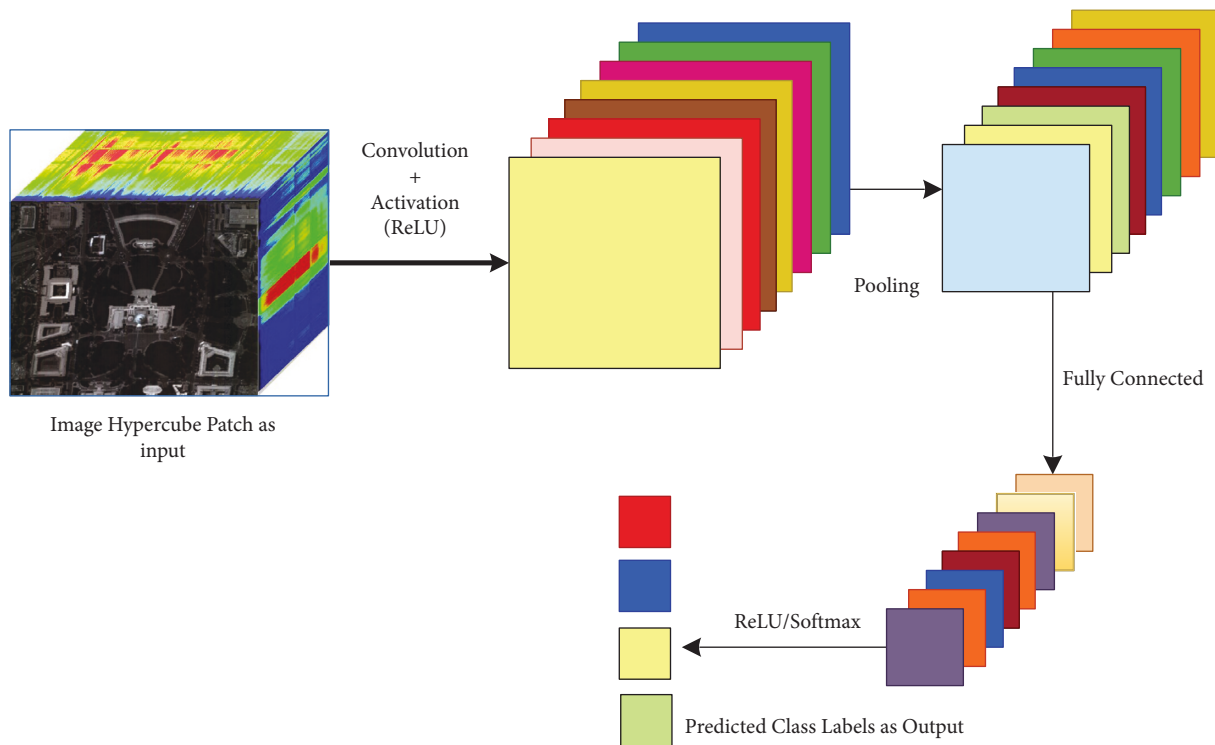
Figure 8: The CNN architecture deploying the layers.

Table 7: Comparison of convolutional layers.

Arguments	Convolution layer	Pooling layer	Fully connected layer
Input	(i) 3D-cube, preceding set of feature maps	(i) 3D-cube, preceding set of feature maps	(i) Flattened-3d-cube, preceding set of feature maps
Parameters	(i) Kernel counts (ii) Kernel size (iii) Activation function (ReLU) (iv) Stride (v) Padding (vi) Type and value of regularization	(i) Stride (ii) Size of window	(i) Number of nodes (ii) Activation function: selected based on the role of the layer. For aggregating info-ReLU. For producing final classification—softmax
Action	(i) Application of filters made of small kernels to extricate features (ii) Learning (iii) One bias per filter (iv) Application of activation function on each feature map value	(i) Reduction of dimensionality (ii) Extraction of the maximum of a region average (iii) Sliding window framework	(i) Aggregate information from final feature maps (ii) Generate final classification
Output	(i) 3D-cube, a 2D-map per filter	(i) 3D-cube, a 2D-map per filter, reduced spatial dimensions	(i) 3D-cube, a 2D-map per filter

TABLE 8: Summary of review of HSI classification using deep learning—CNN.

Year	Method used	Dataset and COA	Research remarks and future scope
2015	Convolutional neural network and multilayer perceptron (CNN-MLP) [120]	Pavia city—99.91%, UP—99.62%, SV—99.53%, IP—98.88%	Far better than SVM, RBF mixed classifiers, the effective convergence rate can be useful for large datasets. Detection of human behavior from hyperspectral video sequences
2016	3D-CNN [121]	IP—98.53%, UP—99.66%, KSC—97.07%	A landmark in terms of quality and overall performance. Mapping performance to be accelerated by postclassification processing
2016	Spectral-spatial feature-based classification (SSFC) [122]	Pavia center—99.87%, UP—96.98%	Highly accurate than other methods. Inclusion of optimal observation scale for improved outcome
2016	CNN-based simple linear iterative clustering (SLIC-CNN) [123]	KSC—100%, UP—99.64, IP—97.24%	Deals with a limited dataset use spectral and local-spatial probabilities as an enhanced estimate in the Bayesian inference
2017	Pixel-pair feature enhanced deep CNN (CNN-PPF) [124]	IP—94.34%, SV—94.8%, UP—96.48%	Overcomes the significant parameter and bulk-data problems of DL, PPFs make the system unique and reliable, and voting strategy makes the more enhanced evaluations in classification
2017	Multiscale 3D deep convolutional neural network (M3D-DCNN) [125]	IP—97.61%, UP—98.49%, SV—97.24%	Outperforms popular methods like RBF-SVM and combinations of CNNs. Removing data limitations and improving the network architecture
2018	2D-CNN, 3D-CNN, recurrent 2D-CNN (R-2D-CNN), and recurrent 3-D-CNN (R-3D-CNN) [126]	IP-99.5%, UP—99.97%, Botswana—99.38%, PaviaC—96.79%, SV—99.8%, KSC—99.85%	R-3D-CNN outperforms all other CNNs mentioned and proves to be very potent in both fast convergence and feature extraction but suffers from the limited sample problem. Applying prior knowledge and transfer learning
2019	3D lightweight convolutional neural network (CNN) (3D-LWNet) [127]	UP—99.4%, IP—98.87%, KSC—98.22%	Provides irrelevance to the sources of data. Architecture is to be improvised by intelligent algorithms
2020	Hybrid spectral CNN (HybridSN) [128]	IP—99.75%, UP—99.98%, SV—100%	Removes the shortfalls of passing over the essential spectral bands and complex, the tedious structure of 2D-CNN and 3D-CNN exclusively and outruns all other contemporary CNN methods superiorly, like SSRN and M-3D-CNN
2020	Heterogeneous TL based on CNN with attention mechanism (HT-CNN-attention) [129]	SV—99%, UP—97.78%, KSC—99.56%, IP—96.99%	Efficient approach regardless of the sample selection strategies chosen
2020	Quantum genetic-optimized SR based CNN (QGASR-CNN) [27]	UP—91.6%, IP—94.1%	With enhanced accuracy, overfitting and "salt-and-pepper" noise are resolved. Improvement of operational performance by the relation between feature mapping and selection of parameters
2020	Rotation-equivariant CNN2D (reCNN2D) [130]	IP—97.78%, UP—98.89, SV—98.18%	Provides robustness and optimal generalization and accuracy without any data augmentation
2020	Spectral-spatial dense connectivity-attention 3D-CNN (SSDANet) [131]	UP—99.97%, IP— 99.29%	Higher accuracy but high computational hazard. Optimization by using other efficient algorithms

$$Y_m = \sum_{k=1}^{r} g\left(i_k . W_m + b_m\right); m = 1, 2, \ldots, d. \quad (23)$$

(2) *Activation*: The convolution layer produces a matrix significantly smaller than the actual image. The matrix is passed through an activation layer (generally rectified linear unit, aka ReLU),

adding nonlinearity that enables the network to train itself through backpropagation.

(3) *Pooling*: It is the method of even more down-sampling and reduction of the matrix size. A filter is applied over the results obtained by the previous layer and chooses a number from each set of values (generally the maximum, the max-

pooling), which allows the network to train much more quickly, concentrating on the most valuable information in each image feature. For an $m \times m$ square window neighbor S with N elements and z_{ij} activation value concerning (i, j) location, the average pooling is formulated as

$$T = \frac{1}{N} \sum_{(i,j) \in S} z_{ij}. \tag{24}$$

(4) *Fully Connected (FC)*: A typical perceptron structure with multilayers. The input is a single-dimensional vector representing the output of the layers above. Its output is a probability list for the various possible labels attached to the image. Classification decision is the mark that receives the highest likelihood. It is mathematically represented with transformation function g, for N samples of inputs with X'' and Y'' being the outputs having W being the weight matrix and b, the bias constant, is as follows:

$$Y'' = \sum_{j=1}^{N} g(WX'' + b). \tag{25}$$

CNN is the most method-in-demand and widely explored model among all DL models. The functional unit of convolutional layers is kernels that expertise in extricating the most relevant and enriched spatial and spectral features from the given dataset through automated filtering by convolution operation [119]. It provides an intense description of the whereabouts of CNNs. The most popular ones are attention-based CNN, ResNet, CapsNet, LeNet, AlexNet, VGG, etc. Some of them are still unexplored yet in classifying HSI. The detailed research work on CNN for dealing with HSI classification is listed in Table 8.

(c) *Recurrent Neural Network (RNN)*: DL is a very efficient approach that follows a sequential framework with a definite timestamp t. "Recurrent" refers to performing the same task for each sequence element, with the output depending on the preceding computations. In other words, they have a "memory" that enfolds information about the calculation so far type of neural network, and the output of a particular recurrent neuron is fed backward as input to the same node, which leads the network to efficiently predict the output, represented in Figure 9, where RNN unrolls, that is, show the complete sequence of the entire network structure neuron by neuron. It consists of the following steps:

(1) $\mathbf{X} = [\ldots, x_{t-1}, x_t, x_{t+1}, \ldots]$ be the input vector, where x_t represents input at timestamp t.
(2) h_t is the "memory of the network," the hidden state at timestamp t. Preliminarily, h_{-1} is initialized to zero vector to calculate the first hidden step. h_t being the current step is calculated based
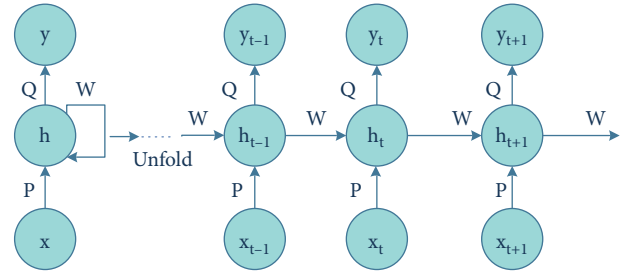

FIGURE 9: The RNN structure with recurrent neurons.

on previously hidden step h_{t-1}, formulated by [132]

$$h_t = f(\mathrm{P}x_t + \mathrm{W}h_{t-1}), \tag{26}$$

where f denotes a function of nonlinearity, that is, tanh or ReLU, and W be the weight vector.
(3) $\mathbf{Y} = [\ldots, y_{t-1}, y_t, y_{t+1}, \ldots]$ be the output vector, where y_t represents input at timestamp t, generally a softmax function: $y_t = \mathrm{softmax}(Q\, h_t)$.
RNN is an efficient deep model with large potential. The recurrence looping structure acquainted with RNN enables it to store relevant information about spatial-spectral relationships between the pixels and neighbors. There are several RNN architectures based on inputs/outputs as stated in [133], and based on LSTM, there are five categories [134]. These variates can be well utilized in collaboration with other DL methods such as MRF and PCA to find their accuracy.
The literature studies based on RNN are cataloged in Table 9.

(d) Deep Belief Network (DBN): DBNs are formed by greedy stacking and training restricted Boltzmann machines (RBMs), an unsupervised learning algorithm based on "contrastive divergence." For neural networks, RBMs suggest taking a probabilistic approach and are thus called stochastic neural networks. Each RBM is made of three parts: a visible unit (input layer), an invisible unit (hidden layer), and a bias unit. The general structure of a DBN is depicted in Figure 10.

For a DBN, the joint distribution of input vector, X with n hidden layers h_n, is defined as [137]

$$P(X, h_1, \ldots, h_n) = \left(\prod_{i=0}^{n-2} P(h_i | h_{i+1}) \right) . P(h_{n-1}, h_n), \tag{27}$$

where $X = h_0$, $P(h_{i-1}, h_i)$ is the conditional distribution of the visible units on the hidden RBM units at level i and $P(h_{n-1}, h_n)$ is the hidden-visible joint distribution in top-level RBM. DBN has two phases: the pretraining phase depicts numerous layers of RBM, and fine-tuning phase is simply a feedforward NN.

TABLE 9: Summary of review of HSI classification using deep learning—RNN.

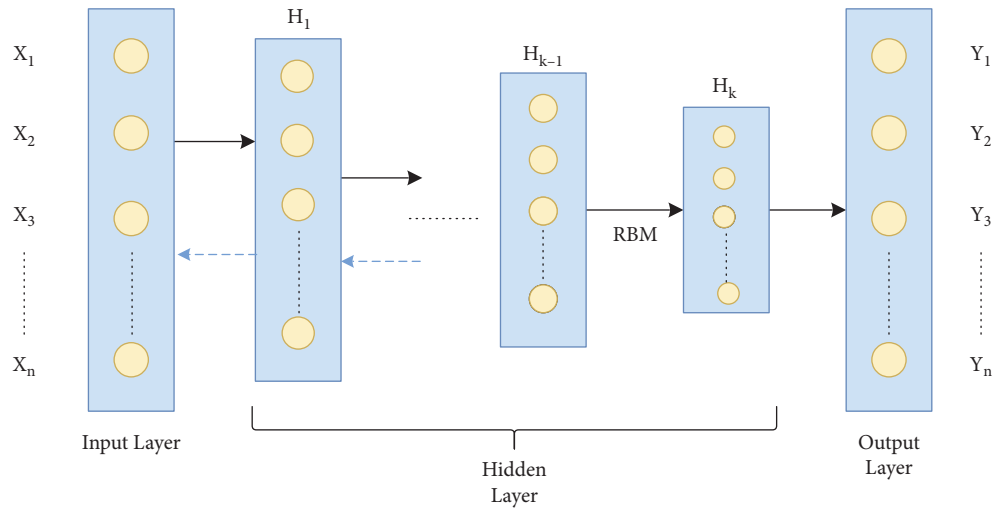| Year	Method used	Dataset and COA	Research remarks and future scope
2017	Gated recurrent unit-based RNN with parametric rectified tanh as activation function (RNN-GRU-pretanh) [132]	UP—88.85%, HU—89.85%, IP—88.63%	An enhanced model that utilizes the intrinsic feature provided by HS pixels with better accuracy than SVM. The study is limited to only spectral features. Incorporation of deep end-to-end convolutional RNN with both spatial-spectral features
2019	Spectral-spatial cascaded recurrent neural network (SSCasRNN) [135]	IP—91.79%, UP—90.30%	Outruns pure RNN and CNN models due to the perfect placement of convolutional and recurrent layers to explore joint information
2020	Geometry-aware deep RNN (Geo-DRNN) [136]	UP—98.05%, IP—97.77%	Due to encoding the complex geometrical structures, the data lack space. Minimization of memory-occupation
2021	2D and 3D spatial attention-driven recurrent feedback convolutional neural network (SARFNN) [28]	IP—99.15%, HU—86.05%	Integrating attention and feedback mechanism with recurrent nets in two layers, 2D and 3D, enables efficient accuracy

FIGURE 10: The detailed DBN structure.

TABLE 10: Summary of review of HSI classification using deep learning—DBN.

Year	Method used	Dataset and COA	Research remarks
2015	Deep belief network and logistic regression (DBN-LR) [137]	IP—95.95%, Pavia City—99.05%	The drawback in training time complexity, it is super-fast testing, and result generating capability outperforms RBF-SVM with EMP
2019	Spectral-adaptive segmented deep belief network (SAS-DBN) [138]	UP—93.15%, HU—98.35%	Capable of addressing the complexities and other subsidiaries of limited samples
2020	Conjugate gradient update-based DBN (CGDBN) [139]	UP—97.31%	Better approach towards stability and convergence of the training model. High time complexity

DBN is the graphical representation that is generative; that is, it creates all distinct outcomes that can be produced for the particular case and learn to disengage a deep hierarchical depiction of the sample training data. DBNs are structurally more capable than RNNs as they lack loops, are pretrained in an unsupervised way, and are computationally eminent for particularly classification problems. Minor modifications or collaborations can improvise DBNs functionally and accuracy. Table 10 depicts a list of works done on DBN.

(e) Generative Adversarial Network (GAN): One of the most recent DL models that are rapidly growing its footsteps in the area of technical research. The GAN model is trained using two kinds of neural networks: the "generative network" or "generator" model that learns to generate new viable samples and the "discriminatory network" or "discriminator," which learns to discriminate generated instances from existing instances. Discriminative algorithms seek to classify the input data, which is given as a collection of certain features; the algorithm maps feature on
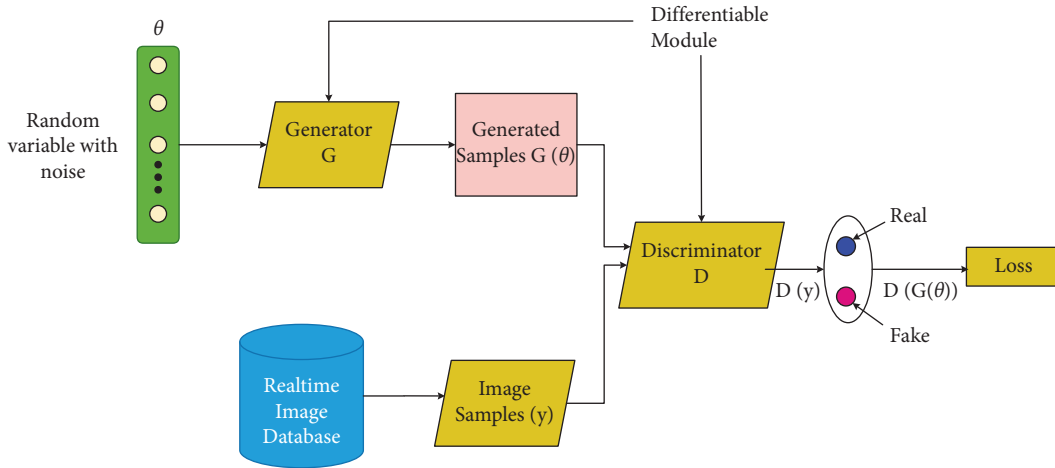
FIGURE 11: The GAN architecture.

labels [140]. In contrast, generative algorithms attempt to construct the input data, which is given with a set of features, and it will not classify it, but it will attempt to create a feature that matches a certain label. The generator tries to get better at deluding the discriminator during the training, and the discriminator tries to grab the counterfeits generated by the generator. Thus, the training procedure is termed adversarial training. The generator and discriminator should be trained against a static opponent, keeping the discriminator constant while training the generator and keeping the generator constant when training the discriminator. That helps to understand the gradients better.

In a GAN model, say D and G denote the discriminator and the generator units that map a noise data space θ to real and original data space x, respectively. $G(\theta)$ denotes the fake output generated by G, and $D(y)$, and $D(G(\theta))$ are D's output for real and fake training samples, respectively. $P_\theta(\theta)$ and $P_d(y)$ represent the input model distribution and original data distribution, respectively, when $\theta \sim P_\theta$ [141] as shown in Figure 11.

The loss function for D:
$$L^{(D)} = \max[\log(D(y)) + \log\ (1 - D(G(\theta)))].$$
(28)

The Loss function for G:
$$L^{(G)} = \min[\log(D(y)) + \log\ (1 - D(G(\theta)))].$$
(29)

Combining equations (28) and (29), the total loss of the entire dataset represented by the min-max value function is given by

$$\min_{G}\max_{D} V(D,G) = \min_{G}\max_{D}\Big(E_y \sim P_d(y)[\log(D(y))] + E_\theta \sim P_\theta(\theta)[\log(1 - D(G(\theta)))]\Big).$$
(30)

GAN is a generative modeling neural network architecture based on the concept of adversarial training that utilizes a model to build new instances that are conceivably derived from an existing sample distribution. Hence, GANs are new favorites for classifying HSIs as they compensate for the lack of data problem and classify the data in a pro manner. There are several types of GANs—conditional GAN, vanilla GAN, deep convolutional GAN (simple type); and Pix2Pix GAN, CycleGAN, StackGAN, and InfoGAN (complex type) [142]. These may be very useful for images like HSIs as they can deal with related issues. The research works based on the GAN are listed in Table 11.

4.7. Transfer Learning (TL). It is the most current hot topic in interactive learning, and there are more to it to be explored. It is an approach where information gained is transferred in one or more source tasks and is used to enhance the learning of a similar target task. TL can be represented diagrammatically by Figure 12 and mathematically shown as follows:

Domain, **D**, is represented as {**X**, $P(X)$}, $X = \{x_1, \ldots, x_n\}$, $x_i \in \mathbf{X}$; **X** denotes the feature space, and $P(X)$ symbolizes the marginal probability of sample data point X [149].

Task **T** is depicted as {**Y**, $P(Y|X)$} = {**Y**, Φ}, $Y = \{y_1, \ldots, y_n\}$, $y_i \in \mathbf{Y}$; **Y** is the label space, Φ is the prognostic objective function, having learned form (feature vector, label) couples, (x_i, y_i); $x_i \in \mathbf{X}$, $y_i \in \mathbf{Y}$, and calculated as the conditional probability.

Also, for every feature vector in **D**, Φ predicts its corresponding label as $\Phi(x_i) = y_i$.

If $\mathbf{D}_S$ and $\mathbf{D}_T$ be the source and target domains, $\mathbf{T}_S$ and $\mathbf{T}_T$ be the source and target tasks, respectively, with $\mathbf{D}_S \neq \mathbf{D}_T$ and $\mathbf{T}_S \neq \mathbf{T}_T$. TL objectifies to learn $P(Y_T|X_T)$, that is, the target conditional probability distribution in $\mathbf{D}_T$ with knowledge obtained from $\mathbf{D}_S$ and $\mathbf{T}_S$.

Traditional learning is segregated and solely based on particular tasks, datasets, and different independent models working on them. No information that can be converted from one model to another is preserved, but on the contrary, TL possesses the human-like capability of transferring

TABLE 11: Summary of review of HSI classification using deep learning—GAN.

| Year	Method used	Dataset and COA	Research remarks and future scope
2018	Hyperspectral 1D generative adversarial networks (HSGAN) [140]	IP—83.53%	Outperforms CNN, KNN, etc.
2018	3D augmented GAN [143]	SV—93.67%, IP—91.1%, KSC—98.12%	Data augmentation solved the problem of overfitting and improved class accuracy
2019	Conditional GAN with conditional variational AE (CGAN-CVAE) [144]	UP—83.85%, DC Mall—89.36%	Semi-supervised and ensemble prediction technique ensures the model's training under limited sample conditions
2020	Semi-supervised variational GAN (SSVGAN) [145]	UP—84.35%, Pavia Center—97.15%, DC Mall—92.21%, Jiamusi—64.76%	Outperforms other GAN variants, that is, CVAEGAN and ACGAN, but it suffers from feature matching, overfitting, and convergence problem Correction through metric learning method
2020	Spectral-spatial GAN-conditional random field (SS-GANCRF) [146]	IP—96.3%, UP—99.31%	Enhanced classification capability Creating an end-to-end training system, graph constraint placed on the convolutional layers
2021	Adaptive weighting feature-fusion generative adversarial network (AWF^2-GAN) [147]	IP—97.53%, UP—98.68%	Exploration of the entire joint feature space and fusion of them, joint loss function, and the central loss gained intraclass sensitivity from local neighboring areas and offered an efficient spatial regularization outcome
2021	Variational generative adversarial network with crossed spatial and spectral interactions (CSSVGAN) [148]	IP—93.61%, UP—99.11%, SV—97%	Increased classification potential by utilizing transformer and GAN
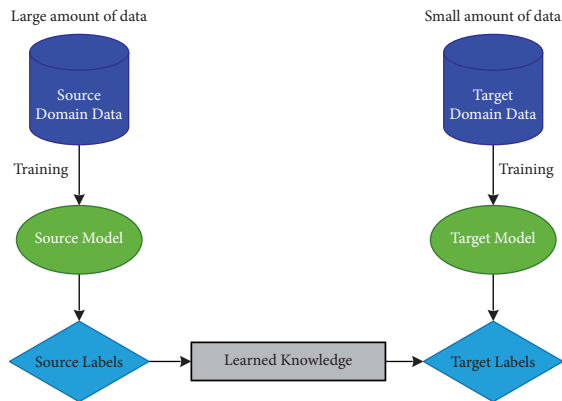
FIGURE 12: The principle of transfer learning.

knowledge; that is, knowledge can be leveraged from priorly trained models to train new models, the process of which is faster, more accurate, and with the limited amount of training data. Table 12 represents a brief detail about the research works on transfer learning.

5. Discussion

Based on the reviewed articles, we can draw the desired inferences that provide answers to the investigative questions mentioned in Section 2 and show the clear motive and benefits of this review.

RI 1: What is the significance of traditional ML and DL for analyzing HSI?

Ans: Hyperspectral data have certain restrictions, as cited in Section 1. Statistical classifiers initially addressed

them, but the operations and analysis became much easier and more accurate after the invention of ML/DL strategies in a machine-dependent way [155, 156]. The general advantages that researchers were provided by the ML/DL algorithms while dealing with HSIs are as follows: (i) easy dealing with high-dimensional data, that is, troubles of Hughes phenomenon removed [115, 125]; (ii) equally manipulative to labeled and unlabeled samples [99, 150]; (iii) precise and the meticulous choice of features [51, 127]; (iv) high-end-precise models to deal with real hypercubes, hence top-notch classification accuracy [119, 154]; v) removes overfitting, noises, and other hurdles to a much greater extent [120, 147]; (vi) embedded spatial-spectral feature extraction and selection units [119, 133]; (vii) mimics human brain to solve multiclass problems [136, 138].

RI 2: How are ML/DL more impactful on HSI than other non-ML strategies?

Ans: The initial discovery of hyperspectral data has suffered due to its limitations. In the preliminary research stage, the scientists followed the traditional methodology for classifying HSIs, that is, preprocessing (if required), extraction, and selection of discriminative characteristics and then ran a classifier on those features to identify the land cover groups. Hence, they emphasized the feature extractor techniques such as PCA [9], ICA [10], and wavelets [13], assisted by some basic random classifiers such as extended morphological profiles [2, 157], NN [158, 159], logistic regression [160], edge-preserving filters [10, 161], density functions/matrices [162], and Bayes law of classification [163, 164]. These classic mathematics-oriented techniques were not enough to deal with such a huge amount of data like HSI, as they were simple in structure and design and easy to

Table 12: Summary of review of HSI classification using transfer learning.

Year	Method used	Dataset and COA	Research remarks and future scope
2018	Deep mapping-based heterogeneous transfer learning model (DLTM) [150]	Washington DC Mall—96.25%	Capable of binary classification Improvisation to multiclass classification
2018	AL with stacked sparse autoencoder (AL-SSAE) [151]	UP—99.48%, center of Pavia—99.8%, SV— 99.45%	Domains, both source, and target possess finely tuned hyperparameters Architectural parameters need to be modified further to enhance the classification accuracy
2020	Heterogeneous TL based on CNN with attention mechanism (HT-CNN-attention) [152]	SV—99%, UP—97.78%, KSC—99.56%, IP—96.99%	Efficient approach regardless of the sample selection strategies chosen
2020	ELM-based ensemble transfer learning (TL-ELM) [26]	UP—98.12%, Pavia center—96.25%	Efficient accuracy and transferability with high training speed Inclusion of SuperPCA and knowledge transfer
2020	Lightweight shuffled group convolutional neural network (abbreviated as SG-CNN) [153]	Botswana—99.67%, HU—99.4%, Washington DC—97.06%	Fine-tuned model as compared to CNN architectures, low computational cost for training Inclusion of more grouped convolutional architectures
2021	Super-pixel pooling convolutional neural network with transfer learning (SP-CNN) [154]	SV—95.99%, UP—93.18%, IP—94.45%	More excellent parameter optimization with more accuracy using a limited number of samples and in a very short period for both training and testing Optimal super-pixel segmentation and merging with different CNN architectures

implement. It also could not predict well enough the multiclass problems, which is very much required for a dataset like HSI, whose land covers belong to multiple classes of regions. Also, these methods were not accurate in feature selection and extraction or dealing with the storage of such bulk data. These reasons made researchers struggle to analyze properly, process, and classify HSIs. On the contrary, the advancements of ML/DL technologies have opened a broad gateway of research that researchers are still exploring and combining with different groupings to address the HSI classification problem in real life, dealing with the limitations mentioned above [26, 131]. The tabular depiction of the advantages and disadvantages of the ML and non-ML strategies applied for HSI classification is shown in Table 13.

RI 3: What are the advantages and challenges faced by the researchers for the chosen ML/DL-based algorithm for HSI classification?

Ans: We added the advantages and challenges of the ML- and DL-based techniques in Table 13.

RI 4: What are the emerging literary works of ML/DL on HSI classification in the year 2021?

Ans: In the ongoing years, 2021 seems to be more promising in terms of technical advancements for the problem concerned. New techniques are emerging, along with hybrid ones, to solve the issue to a whole new level, the methodologies' accuracy to be described. Recent work on MRF with a band-weighted discrete spectral mixture model (MRF-BDSMM) in a Bayesian framework has been proposed in [165], an unsupervised adaptive approach to accommodate heterogeneous noise and find the abundant labeled subpixels to extricate joint features. A collaboration of Kernel-based ELM with PCA, local binary pattern (LBP), and gray-wolf optimization algorithm (PLG) is proposed as novel methodologies. They help reduce huge dimensions, seek global and local-spatial features, and

optimize the KELM parameters to obtain the class labels [166]. A variant of SRC is proposed in [167], dual sparse representation graph-based collaborative propagation (DSRG-CP) that separates spatial and spectral dimensions with the respective graph to improve the labeling scheme limited samples by collaborating the outcomes. AL has been one of the hot topics so far, as it integrates with a Fredholm kernel regularized model (AMKFL) that enables better labeling than manual ones, even for noisy images [168]. It ties with DL with the augmentation of training samples to label the uncertain hypercubes (ADL-UL) accurately [169], facilitates iterative training sample augmentation by expanding the hypercubes and adds discriminative joint features (ITSA-AL-SS) [170], extracts local unique spatial multi-scale characteristics from the super-pixels (MSAL) [171]. A novel idea of attention-based CNNs is proposed in [172, 173], the former (SSAtt-CNN) collides two attention subnetworks—spatial and spectral with CNN as the base, and the latter (FADCNN) is a dense spectral-spatial CNN with feedback attention technique that perfectly poses the band weights for better mining and utilization of dominant features. GAN is one the most exploited methods to date, and [174] proposes the full utilization of shallow features from the unlabeled bands through a multitasking network (MTGAN); in [175], the discriminator is based upon capsule network and convolutional long short-term memory to extricate less visible features and integrates them to build high-profile contextual characteristics (CCAPS-GAN); 1D and 2D CapsGAN together form a dual-channel spectral-spatial fusion capsule GAN (DcCaps-GAN) shown in [176]; and generative adversarial minority oversampling for 3D-hypercubes (3D-Hyper-GAMO) is depicted in [177] that focuses on the minor class features using existing ones to label and classify them properly.

TABLE 13: Comparison between ML and non-ML techniques for HSI classification.

Methods	Advantages	Disadvantages
Classical state-of-art techniques	(i) Simple structure and design (ii) Less time consumption (iii) Easy to implement (iv) Dimension handling skillfully by PCA and ICA (v) Better binary and moderate multiclass classification by kernel and SVM	(i) High space complexity due to the storage of bulk data (ii) Based on empirical identities, hence a tedious workpiece (iii) Feature selection and extraction are not accurate (iv) Suffers from limited labeled sample problem, Hughes phenomenon, and noise
Advanced machine learning techniques	(i) Easy dealing with high-dimensional data, that is, troubles of Hughes phenomenon removed (ii) Equally manipulative to labeled and unlabeled samples (iii) Precise and meticulous choice of features (iv) High-end-precise models to deal with real hypercubes, hence, top-notch classification accuracy (v) Removes overfitting, noises, and other hurdles to a much greater extent (vi) Mimics the human brain to solve multiclass problems	(i) The construction of the model is difficult due to its complex network-alike structure (ii) High time complexity due to training and testing of the huge amount of raw HSI data (iii) Extremely expensive design (iv) Strenuous to implement

RI 5: How are ML- and DL-based hybrid techniques helping scientists in HSI classification?

Ans: Since the dawn of the emergence of HSIs, it has suffered many hurdles in its path of analysis and information extraction. The maximum number of highly correlated bands and the high spatial-spectral features signature by the electromagnetic spectrum embedded in it are always considered a traction matter. Thus, finding an appropriate technology for the classification of such interconnected and hugely confined featured high-dimensional images is a very tedious and strenuous matter. The classification methods chosen so far have been mostly limited to supervised. The requirement of a sufficient number of quality-labeled data and unsupervised, in which the lack of coherence between the spectral clusters and the target regions, causes the failure in obtaining the desired accuracy. A semi-supervised method is needed to overcome such problems as a combination of supervised and unsupervised methods, named the hybrid method. A hybrid method is always advantageous in robustness and flexibility towards the high-dimensional data.

The hybrid methods have the following benefits:

(i) Specifically designed to overcome the limitations and take advantage of the methodologies involved in the concerned hybrid to achieve a deep, rich, and insightful conclusion (general).

(ii) Addressing and resolving multiple issues regarding the handling and analyzing the HSI data, at a time, depending upon the methods that are chosen for mixing/hybridizing [179–183].

(iii) Coherence in time, space, and cost complexities [184–186].

(iv) Better interpretability, quality, effectivity leading to the construction of a more refined framework [180, 182, 183, 187–194].

(v) Deterministic spectral, spatial, and contextual feature extraction, reduction, and selection, and combining them to achieve desired accuracy and performance [182, 183, 187, 188, 195–197].

ML, being a standard versatile technology, can merge with traditional techniques like PCA for its benefit. As stated in [195, 198], PCA is exploited at its best for feature extraction, selection, and reduction to achieve higher accuracy and performance quality. PCA is one of the best pre-processing methods considered to date for improvised spectral dimension reduction [180], proper selection of spectral bands and their multiscale features in a segmented format [181, 199], noise-reduced spectral analysis [27], and feature extraction [130, 196]. PCA, in collaboration with SVM [195, 200], DL for feature reduction and better classification [182, 183], CNN with multiscale feature extraction [188, 189], and sparse tensor technology [190], has highly been appreciated as soulful research. All these recent time collaborations and a special honor to the merging of ICA-DCT with CNN cited in [191] are the evidence that although PCA is categorized under traditional methods, it is supremely relevant for its significant usefulness in handling HSIs.

Some other hybridizations are also explored by researchers, such as SRC with mathematical index of divergence-correlation [192], Gabor-cube filter [193], and ELM [83, 85]; ELM with CNN [86] and TL [26]; AL based on super-pixel profile [201, 202], AL with CNN [203], CapsNet [204], CNN [204, 205], and TL [151, 184]; CNN with attention-aided methodology [172, 173, 185] and GAN [186]; GAN with dynamic neighborhood majority voting mechanism [194, 197], CapsNet [175, 176, 206, 207]; and TL with MRF [70]. These articles depict the highly tenacious performance with literal mitigation of the computational complexities enforced on the raw HSI data to build a strong and enhanced model for achieving higher accuracy than ever.

TABLE 14: The advantages and challenges of the ML- and DL-based techniques for HSI classification.

ML/DL techniques	Advantages	Challenges
Support vector machine	(i) Robust in terms of outliers, Hughes effect, and dimensions as its reduction is not primarily necessary [32, 41, 43] (ii) Supports both supervised, semi-supervised, and unsupervised problems with less overfitting risks [24, 33, 37, 44] (iii) Form of a sigmoid kernel that deals better than the rest of the previous for unlabeled and unstructured HSI datasets [35, 40–42] (iv) The capability of solving the classification problem for both binary and multiclass problems by outperforming several methods [39] (v) Can improve the performance if assisted with other supporting methods [36, 40–42]	(i) It works very well for binary classification but fails for generating accurate classes for multiclass problems [31] (ii) Training time is high for high-class datasets like HSI [31, 32] (iii) Difficulty in fine-tuning the parameters [41, 42] (iv) Complex interpretability [33, 35] (v) Lack of easy generalization to the datasets having multiple classes [33, 35] (vi) Complexity in building the model due to a lack of sufficient labeled samples [31, 32]
Sparse representation and classification	(i) A dictionary with relevant data is used for learning with a minimal number of optimal parameters [45, 46] (ii) Builds precise and powerful classification models with higher interpretability through sparse coding [49, 50, 54] (iii) Proper memory usage in an optimized manner [53, 55, 178] (iv) Reduces the estimated variance between the classes to produce better outcomes [49, 56, 178]	(i) Making the dictionary considers high expense overheads [50] (ii) The dictionary or the coding might cause loss of information [48, 178] (iii) Difficulties in representing such high-profile with higher resolution image data like HSI through the sparse matrix [47, 48]
Markov random field	(i) Works well for a wide range of unstructured problems and no direct dependency between classes and the parameters [67, 69] (ii) Better denoising effect [59] (iii) Robust for both spatial and spectral distributions [62, 64] (iv) Time complexity is low due to the graphical representation of data [63]	(i) Normalization of data might be hectic for high dimension data [63, 70] (ii) Suffers from the lack of training undirected data that might not be possible to represent graphically [61, 62] (iii) Poor interpretability [63, 68]
Extreme learning machines	(i) Less training time and faster learning rate as compared to previous methods [86] (ii) Avoidance for local minima and finishes job in single iteration [83, 87] (iii) Advantageous for overfitting caused due to several bands in HSIs [83] (iv) Builds an enhanced model with better prediction performance at the optimized expense [86] (v) Improved generalization ability, robustness, and controllability [78, 84, 85]	(i) Higher computational hazard [76–80] (ii) The wrong choice of an optimal amount of the hidden layer neurons may cause redundancy in the model and hence affect the classification accuracy [85, 86] (iii) There is plenty of room for advancements in the algorithm to accommodate itself to be compatible for dealing with HSI data [78, 82, 86]

Table 14: Continued.

ML/DL techniques	Advantages	Challenges
Active learning	(i) A very efficient way of learning for both supervised and semi-supervised problems [91, 97, 101, 103] (ii) Ease in segregating the interclass and intraclass features through active query sets [91, 95, 102, 103] (iii) Training speed is comparatively high for not so large-scale data [103] (iv) Knowledge-based solid models can be generated [103] (v) Achieves greater classification accuracies for unlabeled HSIs [95, 102]	(i) Higher computational hazard [76–80] (ii) The wrong choice of an optimal amount of the hidden layer neurons may cause redundancy in the model and hence affect the classification accuracy [85, 86] (iii) There is plenty of room for advancements in the algorithm to accommodate itself to be compatible for dealing with HSI data [78, 82, 86]
Deep learning	(i) Diverse, unstructured, and unlabeled raw HSI datasets are finely processed where preprocessing of the data is not needed [110, 122, 125, 144] (ii) Possesses the capability to address supervised, semi-supervised, and specifically unsupervised learning problems [127, 128, 137] (iii) Expertise in dimension reduction, denoising, feature extraction as embedded properties [27, 114, 124] (iv) Address in an illustrious manner to the issues such as Hughes phenomenon, overfitting, and convergence. [120, 124, 145] (v) Robust and adaptive to new features introduced in the dataset [26, 123, 145] (vi) The hidden layer neurons are proven to be eminent in training the desired model with a highly qualified prior knowledge (DBN, RNN, CNN) [127, 129, 135, 138] (vii) Computational efficiency with high-performance speed (CNN, SAE) [114, 115, 127, 128] (viii) Data augmentation facility (GAN) [143, 145]	(i) Suffers from a lack of a large amount of HSI data, which is practically unavailable [123, 136] (ii) The extreme expense to generate an appropriate model by training a complex data structure like HSIs [114, 139, 148] (iii) Low interpretability [131, 147] (iv) Theoretically not sound, hence incomprehensible where an error occurs and its rectification [122, 124, 145] (v) High time and space complexity and computational hazard [131, 136, 148]
Transfer learning	(i) Works as a combination of different models, be it traditional or latest machine-lefted techniques, that together brings out a highly improved hybrid model [151, 152] (ii) Capable of transferring knowledge from the source domain, that is , a pretrained model to the target domain, that is, a new model to make it more enriched [151, 152] (iii) Greater feature extraction and selection capability [152] (iv) Stable model with highly optimized parameters and hyperparameters [154] (v) High training speed and accuracy with low computational cost [26, 153] (vi) Reduced computational cost and training time complexity [153, 154]	(i) Data overfitting [150] (ii) Complex structure of the model [150, 151] (iii) Less interpretability (iv) Difficulty in implementation

RI 6: What are the latest emerging techniques associated with addressing classifying HSIs?

Ans: The following are the most recent research studies that have enlightened a new path of dealing with the purpose:

(i) *DSVM*: The latest and novel concept incorporates DL facilities with traditional kernel SVM. This combines four deep layers of kernels with SVM being the hidden layer units, namely, exponential and gaussian radial basis function (ERBF and GRBF), neural and polynomial [208]. This approach has outperformed several efficient DL methods with nearly 100% accuracy for IP and UP datasets.

(ii) *Conditional Random Fields (CRFs)*: These are the structured generalization of multinomial logistic regression in the form of graphical models based on a priori continuity considering the neighboring pixels of analogous spectral signatures that possess the same labels. They extensively explore the hidden

spectral-contextual information. In [146], CRF incorporates with semi-supervised GAN whose trained discriminators produce softmax predictions that are guided by dense CRFs graph constraints to improve HSI classification maps. A collaboration between 3D-CNN and CRF has been proposed in [209] to make a deep CRF capable of extracting the semantic correlations between patches of hypercubes by CNN's unary and pairwise potential functions. A semi-supervised approach is depicted in [210], embedding subspace learning and 3D convolutional autoencoder to remove redundancy in joint features and obtain class sets using an iterative algorithm. In [211], CRF with Gaussian edge potentials associated with deep metric learning (DML) classifies HSI data pixelwise using the geographical distances between pixels and the Euclidean distances between the features. A novel framework using HSI feature learning network (HSINet) with CRF is proposed [212] that is a trainable end-to-end DL model with backpropagation that extracts joint features, edges, and colors based on subpixel, pixel, and super-pixels. In [213], a decision fusion model including CRF and MRF is built based on sparse unmixing and soft classifiers output.

(iii) *Random Forest (RF)*: It is an efficient algorithm that ensembles regression and classification tree. It enables the HSI classification model to be noise-tolerant, inherent in the multiclass division, robustness in parallelism, and speed. In [214], RF is compared to the DL algorithm, which outshined the classification accuracy. A new framework of cascaded RF is shown in [215] that uses the boosting strategy to generate and train base classifiers and Hierarchical Random Subspace Method to select features and suitable base classifiers based on the diversity of the features. A novel collaboration of semi-supervised learning and AL and RF is featured in [216], where the queries based on spatial information are fed to AL, and then, the labeled samples are classified by RF through semi-supervision. [217, 218] depicts a deep cube CNN model that extracts pixelwise joint features and is classified by RF.

(iv) *Graph Convolutional Network (GCN)*: A descendent of CNN, a structure designed to generalize and convert the convolution data to graph data. It consists of three steps feature aggregation, feature transformation, and classification. Being an expert in graphical modeling considers the spatial interrelations between the classes at its best. In [219], the different unique features collected from CNN and GCN are fused additive, elementwise, and concatenated way. A new framework of globally consistent GCN is introduced in [220], which first generates a spatial-spectral local optimized graph whose global high-order neighbors obtain the enriched contextual information employing the graph topological

consistent connectivity; at last, those global features determine the classes. [221] shows the concept of a dual GCN network, which works with a limited number of training samples, where first extricates all the significant features and second learns label distribution. A novel idea of deep attention GCN is introduced in [222] based on similarity measurement criteria between the mixed measurement of a kernel-spectral angle mapper and spectral information divergence to accumulate analogous spectra. [223] emerges as a collaboration between CNN and GCN to extract pixel and super-pixelwise joint features by learning small-scale regular regions and large-scale irregular regions.

6. Conclusion

This article depicts the various technologies and procedures used for HSI classification since the dawn of its invention to date. There are many barriers to dealing with such high-band data as HSI mentioned above. Despite that, many researchers have taken their interest in this field to improvise the existing techniques or even invent new ones throughout the last decade. As per the considerable improvement in technologies and the introduction of ML into the classification issues of HSI, it has become more accurate than traditional and contemporary state-of-art methodologies. As a result, DL has emerged as the most eminent work tool for HSI classification for the last half of this decade. The more the researchers focused on this, the more they explored the remote sensing and space imagery features.

This review article bears the individual information for every method and their submethods about their performance, research gaps, and achievements. In addition, it appends a novel research methodology that makes this work more distinctive than others. After going through each methodology's minute details, the most significant inferences have been drawn, which add further novelty to our work. Also, it shows a path of choosing an appropriate technique and its alternatives for future researchers, hence alleviating its creativity and uniqueness, above all other contemporary review works on this subject. Also, it provides the details of the most recent research scenario on HSI classification and some of the currently developed techniques that might be acutely useful in several future research. Our study holds the uniqueness and the novelty regarding several aspects, such as the following: (1) it includes the research works carried out in the last decade, that is, 2010–2020, and the most recent papers of the previous year, i.e., 2021, and we have mentioned it in Section 3; (2) the number of papers referred here is above 200, outnumbering other review papers; (3) the review is carried out by selecting the most appropriate papers solely dedicated to our subject of interest, that is, machine learning techniques serving the purpose of hyperspectral image classification. Then, the findings from those works of literature are systematically arranged in the tabular format (Tables 1–12); (4) the

objective behind this review work is expressed by RQ 1–6. Also, they provide a clear view of the recent technological advances and applications that the researchers are developing in recent times; (5) Table 14 provides an explicit idea of the pros and cons of each ML technique described in this manuscript when applied for classifying hyperspectral images, which will help the researchers in their future research; and (6) the researcher who wishes to write a literature review can follow our proposed methodology that depicts the flow of work in a methodical way. [224].

7. Limitations of Present Work and Its Future Scope

The study has some limitations: (i) we have used fewer keywords in the current research (ii) we only focused on seven popular ML techniques; (iii) we briefly explain the emerging methodologies; and (iv) the experimental details are not fully discussed.

As a future proposition, we would like to explore more keywords, more techniques, and more studies that offer a better understanding of other learning methods, both traditional and contemporary. In addition, there are several instances of hybrid strategies along with some more eminent and latest ML/DL techniques that we shall look forward to exploring in both qualitative and quantitative manner.

Acronym

HS: Hyperspectral
HSI: Hyperspectral image
GIS: Geographic Information System
PCA: Principal component analysis
ICA: Independent component analysis
SVM: Support vector machine
SR: Sparse representation
SRC: Sparse representation and classification
MRF: Markov random field
HMRF: Hidden Markov random field
ELM: Extreme learning machine
AL: Active learning
HU: University of Houston
TL: Transfer learning
DL: Deep learning
AE: Autoencoders
SAE: Stacked autoencoders
CNN: Convolutional neural network
RNN: Recurrent neural network
DBN: Deep belief network
GAN: Generative adversarial network
IP: Indian pines
KSC: Kennedy space center
SV: Salinas valley
UP: University of Pavia.

Data Availability

Publicly available data are used in this study.

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

References

[1] M. J. Khan, H. S. Khan, A. Yousaf, K. Khurshid, and A. Abbas, "Modern trends in hyperspectral image analysis: a review," *IEEE Access*, vol. 6, pp. 14118–14129, 2018.

[2] N. Falco, J. A. Benediktsson, and L. Bruzzone, "Spectral and spatial classification of hyperspectral images based on ICA and reduced morphological attribute profiles," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 11, pp. 6223–6240, 2015.

[3] T. Adão, J. Hruška, L. Pádua et al., "Hyperspectral imaging: a review on UAV-based sensors, data processing and applications for agriculture and forestry," *Remote Sensing*, vol. 9, p. 1110, 2017.

[4] C. Northcutt, L. Jiang, and I. Chuang, "Confident learning: estimating uncertainty in dataset labels," https://www.jair.org/index.php/jair/article/view/12125.

[5] Z. Xu, D. Lu, Y. Wang et al., "Noisy labels are treasure: mean-teacher-assisted confident learning for hepatic vessel segmentation," in *Proceedings of the Medical Image Computing and Computer Assisted Intervention – MICCAI 2021*, Springer, Strasbourg, France, 27 September 2021, https://link.springer.com/chapter/10.1007%2F978-3-030-87193-2_1.

[6] C. G. Northcutt, T. Wu, and I. L. Chuang, "Learning with confident examples: rank pruning for robust classification with noisy labels," https://arxiv.org/abs/1705.01936.

[7] P. Ghamisi, N. Yokoya, J. Li et al., "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, Dec. 2017.

[8] T. Han and D. G. Goodenough, "Investigation of nonlinearity in hyperspectral imagery using surrogate data methods," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, no. 10, pp. 2840–2847, Oct. 2008.

[9] B. A. Beirami and M. Mokhtarzade, "Band grouping SuperPCA for feature extraction and extended morphological profile production from hyperspectral images," *IEEE Geoscience and Remote Sensing Letters (Early Access)*, vol. 17, pp. 1–5, 2020.

[10] J. Xia, L. Bombrun, T. Adalı, Y. Berthoumieu, and C. Germain, "Spectral–spatial classification of hyperspectral images using ICA and edge-preserving filter via an ensemble strategy," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 8, pp. 4971–4982, 2016.

[11] M. Imani and H. Ghassemian, "Principal component discriminant analysis for feature extraction and classification of hyperspectral images," in *Proceedings of the 2014 Iranian Conference on Intelligent Systems (ICIS)*, IEEE, Bam, Iran, 4 February 2014.

[12] W. Li, S. Prasad, J. E. Fowler, and L. M. Bruce, "Locality-preserving discriminant analysis in kernel-induced feature spaces for hyperspectral image classification," *IEEE Geoscience and Remote Sensing Letters*, vol. 8, no. 5, pp. 894–898, 2011.

[13] X. Cao, J. Yao, X. Fu, H. Bi, and D. Hong, "An enhanced 3-D discrete wavelet transform for hyperspectral image classification," *IEEE Geoscience and Remote Sensing Letters (Early Access)*, vol. 18, pp. 1–5, 2020.

[14] J. Peng, H. Chen, Y. Zhou, and L. Li, "Ideal regularized composite kernel for hyperspectral image classification," *Ieee Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 10, no. 4, pp. 1563–1574, 2017.

[15] J. Li, P. R. Marpu, A. Plaza, J. M. Bioucas-Dias, and J. A. Benediktsson, "Generalized composite kernel framework for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 51, no. 9, pp. 4816–4829, 2013.

[16] J. Liu, Z. Wu, J. Li, A. Plaza, and Y. Yuan, "Probabilistic-kernel collaborative representation for spatial–spectral hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 4, pp. 2371–2384, 2016.

[17] M. S. Kumar, V. Keerthi, R. N. Anjnai, M. M. Sarma, and V. Bothale, "Evalution of machine learning methods for hyperspectral image classification," in *Proceedings of the 2020 IEEE India Geoscience and Remote Sensing Symposium (InGARSS)*, pp. 225–228, IEEE, Ahmedabad, India, 1 December 2020.

[18] "Hyperspectral remote sensing scenes," http://www.ehu.eus/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes.

[19] M. E. Paoletti, J. M. 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, December, 2019.

[20] C. L. Chowdhary, P. V. Patel, K. J. Kathrotia, M. Attique, K. Perumal, and M. F. Ijaz, "Analytical study of hybrid techniques for image encryption and decryption," *Sensors*, vol. 20, no. 18, p. 5162, 2020.

[21] 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.

[22] Y. Quan, X. Zhong, W. Feng, J. C.-W. Chan, Q. Li, and M. Xing, "SMOTE-based weighted deep rotation forest for the imbalanced hyperspectral data classification," *Remote Sensing*, vol. 13, no. 3, p. 464, 2021.

[23] G. Hughes, "On the mean accuracy of statistical pattern recognizers," *IEEE Transactions on Information Theory*, vol. 14, no. 1, pp. 55–63, January 1968.

[24] D. K. Pathak and S. K. Kalita, "Spectral spatial feature based classification of hyperspectral image using support vector machine," in *Proceedings of the 2019 6th International Conference on Signal Processing and Integrated Networks (SPIN), Date of Conference*, IEEE, Noida, India, 7 March 2019.

[25] L. Zhou and L. Ma, "Extreme learning machine-based heterogeneous domain adaptation for classification of hyperspectral images," *IEEE Geoscience and Remote Sensing Letters*, vol. 16, no. 11, pp. 1781–1785, 2019.

[26] 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.

[27] H. Chen, F. Miao, and X. Shen, "Hyperspectral remote sensing image classification with CNN based on quantum genetic-optimized sparse representation," *IEEE Access*, vol. 8, pp. 99900–99909.

[28] H. C. Li, S. S. Li, W. S. Hu, J. H. Feng, W. W. Sun, and Q. Du, "Recurrent feedback convolutional neural network for hyperspectral image classification," *IEEE Geoscience and Remote Sensing Letters*, vol. 19, 2021.

[29] S. Mian Qaisar, "Signal-piloted processing and machine learning based efficient power quality disturbances recognition," *PLoS One*, vol. 16, no. 5, Article ID e0252104, 2021.

[30] F. Chamasemani and Y. P. Singh, "Multi-class support vector machine (SVM) classifiers -- an application in hypothyroid detection and classification," in *Proceedings of the The 2011 Sixth International Conference on Bio-Inspired Computing*, pp. 351–356, IEEE, Penang, Malaysia, 27 Setember 2011.

[31] J. Zhang, Y. Zhang, and T. Zhou, "Classification of hyperspectral data using support vector machine," in *Proceedings of the 2001 International Conference on Image Processing (Cat. No.01CH37205)*, IEEE, Thessaloniki, Greece, 7 October 2001.

[32] G. Camps-Valls, L. Gomez-Chova, J. Calpe-Maravilla et al., "Robust support vector method for hyperspectral data classification and knowledge discovery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 42, no. 7, pp. 1530–1542, July 2004.

[33] F. Melgani and L. Bruzzone, "Classification of hyperspectral remote sensing images with support vector ma-chines," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 42, no. 8, pp. 1778–1790, 2004.

[34] B. Borasca, L. Bruzzone, L. Carlin, and M. Zusi, "A fuzzy-input fuzzy-output SVM technique for classification of hyperspectral remote sensing images," in *Proceedings of the 7th Nordic Signal Processing Symposium - NORSIG 2006*, IEEE, Reykjavik, Iceland, 7 June 2006.

[35] M. Fauvel, J. A. Benediktsson, J. Chanussot, and J. R. Sveinsson, "Spectral and spatial classification of hyperspectral data using SVMs and morphological profiles," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, no. 11, pp. 3804–3814, 2008.

[36] S. Ding and L. Chen, "Classification of hyperspectral remote sensing images with support vector machines and particle swarm optimization," in *Proceedings of the International Conference on Information Engineering and Computer Science*, 19 December 2009.

[37] P. Du, K. Tan, and X. Xing, "Wavelet SVM in Reproducing Kernel Hilbert Space for hyperspectral remote sensing image classification," *Optics Communications*, vol. 283, no. 24, pp. 4978–4984, 2010.

[38] F. A. Mianji and Y. Zhang, "Semisupervised support vector machine classification for hyperspectral imagery," in *Proceedings of the 2011 International Conference on Communications and Signal Processing*, 10 February 2011.

[39] S. Moustakidis, G. Mallinis, N. Koutsias, J. B. Theocharis, and V. Petridis, "SVM-based fuzzy decision trees for classification of high spatial resolution remote sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 1, pp. 149–169, Jan. 2012.

[40] Z. Shao, L. Zhang, X. Zhou, and L. Ding, "A novel hierarchical semisupervised SVM for classification of hyperspectral images," *IEEE Geoscience and Remote Sensing Letters*, vol. 11, no. 9, pp. 1609–1613, 2014.

[41] B. Kuo, H. Ho, C. Li, C. Hung, and J. Taur, "A kernel-based feature selection method for SVM with RBF kernel for hyperspectral image classification," *Ieee Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, no. 1, pp. 317–326, Jan. 2014.

[42] J. Peng, Y. Zhou, and C. L. P. Chen, "Region-kernel-based support vector machines for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 9, pp. 4810–4824, 2015.

[43] H. Yu, L. Gao, W. Liao, B. Zhang, A. Pižurica, and W. Philips, "Multiscale super-pixel-level subspace-based support vector machines for hyperspectral image classification," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 11, pp. 2142–2146, 2017.

[44] C. Zhang, M. Han, and M. Xu, "Multi-feature classification of hyperspectral image via probabilistic SVM and guided filter," in *Proceedings of the 2018 International Joint Conference on Neural Networks (IJCNN)*, IEEE, Rio de Janeiro, Brazil, 8 July 2018.

[45] J. Liu, Z. Wu, Z. Wei, L. Xiao, and L. Sun, "Spatial-spectral kernel sparse representation for hyperspectral image classification," *Ieee Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 6, no. 6, pp. 2462–2471, Dec. 2013.

[46] L. Fang, S. Li, X. Kang, and J. A. Benediktsson, "Spectral–spatial hyperspectral image classification via multiscale Adaptive sparse representation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 12, pp. 7738–7749, Dec. 2014.

[47] P. Du, Z. Xue, J. Li, and A. Plaza, "Learning discriminative sparse representations for hyperspectral image classification," *IEEE Journal of Selected Topics in Signal Processing*, vol. 9, no. 6, pp. 1089–1104, Sept. 2015.

[48] W. Fu, S. Li, L. Fang, X. Kang, and J. A. Benediktsson, "Hyperspectral image classification via shape-adaptive joint sparse representation," *Ieee Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 9, no. 2, pp. 556–567, Feb. 2016.

[49] L. Fang, C. Wang, S. Li, and J. A. Benediktsson, "Hyperspectral image classification via multiple-feature-based adaptive sparse representation," *IEEE Transactions on Instrumentation and Measurement*, vol. 66, no. 7, pp. 1646–1657, July 2017.

[50] B. Tu, S. Huang, L. Fang, G. Zhang, J. Wang, and B. Zheng, "Hyperspectral image classification via weighted joint nearest neighbor and sparse representation," *Ieee Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 11, no. 11, pp. 4063–4075, 2018.

[51] W. Yang, J. Peng, W. Sun, and Q. Du, "Log-euclidean kernel-based joint sparse representation for hyperspectral image classification," *Ieee Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 12, no. 12, pp. 5023–5034, Dec. 2019.

[52] T. Dundar and T. Ince, "Sparse representation-based hyperspectral image classification using multiscale super-pixels and guided filter," *IEEE Geoscience and Remote Sensing Letters*, vol. 16, no. 2, pp. 246–250, Feb. 2019.

[53] J. Peng, W. Sun, and Q. Du, "Self-paced joint sparse representation for the classification of hyperspectral images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 2, pp. 1183–1194, Feb. 2019.

[54] J. Peng, L. Li, and Y. Y. Tang, "Maximum likelihood estimation-based joint sparse representation for the classification of hyperspectral remote sensing images," *IEEE*

[55] H. Yu, L. Gao, W. Liao et al., "Global spatial and local spectral similarity-based manifold learning group sparse representation for hyperspectral imagery classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 5, pp. 3043–3056, May 2020.

[56] F. Luo, L. Zhang, X. Zhou, T. Guo, Y. Cheng, and T. Yin, "Sparse-adaptive hypergraph discriminant analysis for hyperspectral image classification," *IEEE Geoscience and Remote Sensing Letters*, vol. 17, no. 6, pp. 1082–1086, June 2020.

[57] "Markov random field," https://en.wikipedia.org/wiki/Markov_random_field.

[58] G. Altalib and E. Ahmed, "Land cover classification using hidden Markov models," *International Journal of Computer Networks and Communications Security*, vol. 1, pp. 165–172, 2013.

[59] 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.

[60] P. Ghamisi, J. A. Benediktsson, and M. O. Ulfarsson, "Spectral–spatial classification of hyperspectral images based on hidden Markov random fields," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 5, pp. 2565–2574, May 2014.

[61] L. Xu and J. Li, "Bayesian classification of hyperspectral imagery based on probabilistic sparse representation and Markov random field," *IEEE Geoscience and Remote Sensing Letters*, vol. 11, no. 4, pp. 823–827, April 2014.

[62] W. Li, S. Prasad, and J. E. Fowler, "Hyperspectral image classification using Gaussian mixture models and Markov random fields," *IEEE Geoscience and Remote Sensing Letters*, vol. 11, no. 1, pp. 153–157, Jan. 2014.

[63] L. Sun, Z. Wu, J. Liu, L. Xiao, and Z. Wei, "Supervised spectral–spatial hyperspectral image classification with weighted Markov random fields," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 3, pp. 1490–1503, March 2015.

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

[65] M. Golipour, H. Ghassemian, and F. Mirzapour, "Integrating hierarchical segmentation maps with MRF prior for classification of hyperspectral images in a bayesian framework," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 2, pp. 805–816, Feb. 2016.

[66] E. K. Ghasrodashti, M. S. Helfroush, and H. Danyali, "Sparse-based classification of hyper-spectral images using extended hidden Markov random fields," *Ieee Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 11, no. 11, pp. 4101–4112, 2018.

[67] Y. Fang, L. Xu, J. Peng, H. Yang, A. Wong, and D. A. Clausi, "Unsupervised bayesian classification of a hyperspectral image based on the spectral mixture model and Markov random field," *Ieee Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 11, no. 9, pp. 3325–3337, 2018.

[68] C. Pan, X. Gao, Y. Wang, and J. Li, "Markov random fields integrating adaptive interclass-pair penalty and spectral similarity for hyperspectral image classification," *IEEE*

Transactions on Geoscience and Remote Sensing, vol. 57, no. 5, pp. 2520–2534, May 2019.

[69] X. Cao, X. Wang, D. Wang, J. Zhao, and L. Jiao, "Spectral–spatial hyperspectral image classification using cascaded Markov random fields," *Ieee Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 12, no. 12, pp. 4861–4872, Dec. 2019.

[70] X. Jiang, Y. Zhang, Y. Li, S. Li, and Y. Zhang, "Hyperspectral image classification with transfer learning and Markov random fields," *IEEE Geoscience and Remote Sensing Letters*, vol. 17, no. 3, pp. 544–548, March 2020.

[71] X. Jiang, Y. Zhang, W. Liu et al., "Hyperspectral image classification with CapsNet and Markov random fields," *IEEE Access*, vol. 8, pp. 191956–191968.

[72] A. Samat, P. Du, S. Liu, J. Li, and L. Cheng, "E2LMs: ensemble extreme learning machines for hyperspectral image classification," *Ieee Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, no. 4, pp. 1060–1069, April 2014.

[73] S. Ding, H. Zhao, Y. Zhang, X. Xu, and R. Nie, "Extreme learning machine: algorithm, theory and applications," *Artificial Intelligence Review*, vol. 44, no. 1, pp. 103–115, 2015.

[74] "A multiple hidden layers extreme learning machine method and its application," https://www.hindawi.com/journals/mpe/2017/4670187/.

[75] Y. Zhou, J. Peng, and C. L. P. Chen, "Extreme learning machine with composite kernels for hyperspectral image classification," *Ieee Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 8, no. 6, pp. 2351–2360, June 2015.

[76] W. Li, C. Chen, H. Su, and Q. Du, "Local binary patterns and extreme learning machine for hyperspectral imagery classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 7, pp. 3681–3693, July 2015.

[77] Q. Lv, X. Niu, Y. Dou, J. Xu, and Y. Lei, "Classification of hyperspectral remote sensing image using hierarchical local-receptive-field-based extreme learning machine," *IEEE Geoscience and Remote Sensing Letters*, vol. 13, no. 3, pp. 434–438, March 2016.

[78] H. Su, Y. Cai, and Q. Du, "Firefly-algorithm-inspired framework with band selection and extreme learning machine for hyperspectral image classification," *Ieee Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 10, no. 1, pp. 309–320, Jan. 2017.

[79] Y. Shen, J. Chen, and L. Xiao, "Supervised classification of hyperspectral images using local-receptive-fields-based kernel extreme learning machine," in *Proceedings of the 2017 IEEE International Conference on Image Processing (ICIP)*, IEEE, Beijing, China, 17–20 September 2017.

[80] J. Ku and B. Zheng, "Distributed extreme learning machine with kernels based on MapReduce for spectral-spatial classification of hyperspectral image," in *Proceedings of the 2017 IEEE International Conference on Computational Science and Engineering (CSE) and IEEE International Conference on Embedded and Ubiquitous Computing (EUC)*, IEEE, Guangzhou, China, 21 July 2017.

[81] F. Cao, Z. Yang, M. Jiang, W. Chen, Q. Ye, and W. Ling, "Spectral-spatial classification of hyperspectral image using extreme learning machine and loopy Belief propagation," in *Proceedings of the 2017 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData)*, IEEE, Exeter, UK, 21 June 2017.

[82] W. Shang, Z. Wu, Y. Xu, Y. Zhang, and Z. Wei, "Hyperspectral supervised classification using mean filtering based kernel extreme learning machine," in *Proceedings of the 2018 Fifth International Workshop on Earth Observation and Remote Sensing Applications (EORSA)*, IEEE, Xi'an, China, 18 June 2018.

[83] F. Cao, Z. Yang, J. Ren et al., "Sparse representation-based augmented multinomial logistic extreme learning machine with weighted composite features for spectral–spatial classification of hyperspectral images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 11, pp. 6263–6279, 2018.

[84] M. Jiang, F. Cao, and Y. Lu, "Extreme learning machine with enhanced composite feature for spectral-spatial hyperspectral image classification," *IEEE Access*, vol. 6, pp. 22645–22654.

[85] F. Cao, Z. Yang, J. Ren, W. Chen, G. Han, and Y. Shen, "Local block multilayer sparse extreme learning machine for effective feature extraction and classification of hyperspectral images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 8, pp. 5580–5594, 2019.

[86] Y. Shen, L. Xiao, J. Chen, and D. Pan, "A spectral-spatial domain-specific convolutional deep extreme learning machine for supervised hyperspectral image classification," *IEEE Access*, vol. 7, pp. 132240–132252.

[87] Y. Yin and L. Wei, "Hyperspectral image classification using comprehensive evaluation model of extreme learning machine based on cumulative variation weights," *IEEE Access*, vol. 8, p. 187991, 188003.

[88] "Introduction to active learning," https://towardsdatascience.com/introduction-to-active-learning-117e0740d7cc.

[89] "Active learning machine learning: what it is and how it works," https://algorithmia.com/blog/active-learning-machine-learning.

[90] S. Rajan, J. Ghosh, and M. M. Crawford, "An active learning approach to hyperspectral data classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, no. 4, pp. 1231–1242, April 2008.

[91] J. Li, J. M. Bioucas-Dias, and A. Plaza, "Semisupervised hyperspectral image segmentation using multinomial logistic regression with active learning," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 11, pp. 4085–4098, 2010.

[92] J. Li, J. M. Bioucas-Dias, and A. Plaza, "Spectral–spatial classification of hyperspectral data using loopy Belief propagation and active learning," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 51, no. 2, pp. 844–856, Feb. 2013.

[93] S. Sun, P. Zhong, H. Xiao, and R. Wang, "An MRF model-based active learning framework for the spectral-spatial classification of hyperspectral imagery," *IEEE Journal of Selected Topics in Signal Processing*, vol. 9, no. 6, pp. 1074–1088, Sept. 2015.

[94] S. Sun, P. Zhong, H. Xiao, and R. Wang, "Active learning with Gaussian process classifier for hyper-spectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 4, pp. 1746–1760, April 2015.

[95] Z. Zhang, E. Pasolli, M. M. Crawford, and J. C. Tilton, "An active learning framework for hyperspectral image classification using hierarchical segmentation," *Ieee Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 9, no. 2, pp. 640–654, Feb. 2016.

[96] X. Zhou, S. Prasad, and M. M. Crawford, "Wavelet-domain multiview active learning for spatial-spectral hyperspectral

image classification," *Ieee Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 9, no. 9, pp. 4047–4059, 2016.

[97] Z. Wang, B. Du, L. Zhang, L. Zhang, and X. Jia, "A novel semisupervised active-learning algorithm for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 6, pp. 3071–3083, June 2017.

[98] 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, Dec. 2017.

[99] 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, Jan. 2018.

[100] 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, Feb. 2018.

[101] C. Liu, J. Li, and L. He, "Super-pixel-Based semisupervised active learning for hyperspectral image classification," *Ieee Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 12, no. 1, pp. 357–370, Jan. 2019.

[102] 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, April 2020.

[103] C. Mu, J. Liu, Y. Liu, and Y. Liu, "Hyperspectral image classification based on active learning and spectral-spatial feature fusion using spatial coordinates," *IEEE Access*, vol. 8, pp. 6768–6781, 03 January 2020.

[104] 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.

[105] "What is deep learning and how does it work," https://towardsdatascience.com/what-is-deep-learning-and-how-does-it-work-f7d02aa9d477.

[106] A. Subasi and S. Mian Qaisar, "The ensemble machine learning-based classification of motor imagery tasks in brain-computer interface," *Journal of Healthcare Engineering*, vol. 2021, Article ID 1970769, 12 pages, 2021.

[107] B. Alsinglawi, O. Alshari, M. Alorjani et al., "An explainable machine learning framework for lung cancer hospital length of stay prediction," *Scientific Reports*, vol. 12, no. 1, pp. 1–10, 2022.

[108] B. Alsinglawi, F. Alnajjar, O. Mubin, M. Novoa, O. Karajeh, and O. Darwish, "Benchmarking predictive models in electronic health records: sepsis length of stay prediction," in *Proceedings of the International Conference on Advanced Information Networking and Applications*, pp. 258–267, Springer, Caserta, Italy, 15 April 2020, https://link.springer.com/chapter/10.1007/978-3-030-44041-1_24.

[109] P. N. Srinivasu, J. G. SivaSai, M. F. Ijaz, A. K. Bhoi, W. Kim, and J. J. Kang, "Classification of skin disease using deep learning neural networks with MobileNet V2," *Sensors*, vol. 21, no. 8, p. 2852.

[110] Z. Lin, Y. Chen, X. Zhao, and G. Wang, "Spectral-spatial classification of hyperspectral image using autoencoders," in *Proceedings of the 2013 9th International Conference on Information, Communications & Signal Processing*, IEEE, Tainan, 10 December 2013.

[111] G. Liu, H. Bao, and B. Han, "A stacked autoencoder-based deep neural network for achieving gearbox fault diagnosis," *Advancements in Mathematical Methods for Pattern Recognition and its Applications*, vol. 2018, Article ID 5105709, 10 pages, 2018.

[112] A. Tutorial, "A beginner's guide to autoencoders," https://www.edureka.co/blog/autoencoders-tutorial/.

[113] 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, June 2014.

[114] X. Ma, H. Wang, and J. Geng, "Spectral–spatial classification of hyperspectral image based on deep auto-encoder," *Ieee Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 9, no. 9, pp. 4073–4085, Sept. 2016.

[115] 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, July 2019.

[116] H. Madani and K. McIsaac, "Distance transform-based spectral-spatial feature vector for hyperspectral image classification with stacked autoencoder," *Remote Sensing*, vol. 13, p. 1732, 2021.

[117] "Layers of a convolutional neural network," https://wiki.tum.de/display/lfdv/Layers+of+a+Convolutional+Neural+Network.

[118] B. C. N. N. Architecture, "Explaining 5 layers of convolutional neural network," https://www.upgrad.com/blog/basic-cnn-architecture/.

[119] "A survey of the recent architectures of deep convolutional neural networks," https://arxiv.org/ftp/arxiv/papers/1901/1901.06032.pdf.

[120] K. Makantasis, K. Karantzalos, A. Doulamis, and N. Doulamis, "Deep supervised learning for hyperspectral data classification through convolutional neural networks," in *Proceedings of the 2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Date of Conference*, IEEE, Milan, Italy, 26 July 2015.

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

[122] W. Zhao and S. Du, "Spectral–spatial feature extraction for hyperspectral image classification: a dimension reduction and deep learning approach," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 8, pp. 4544–4554, 2016.

[123] J. Cao, Z. Chen, and B. Wang, "Deep convolutional networks with super-pixel segmentation for hyperspectral image classification," in *Proceedings of the 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 10 July 2016.

[124] W. Li, G. Wu, F. Zhang, and Q. Du, "Hyperspectral image classification using deep pixel-pair features," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 2, pp. 844–853, Feb. 2017.

[125] M. He, B. Li, and H. Chen, "Multi-scale 3d deep convolutional neural network for hyper-spectral image classification," in *Proceedings of the 2017 IEEE International Conference on Image Processing (ICIP)*, IEEE, Beijing, China, 17 September 2017.

[126] X. Yang, Y. Ye, X. Li, R. Y. K. 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.

[127] 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.

[128] 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, Feb. 2020.

[129] W. Hu, H. Li, L. Pan, W. Li, R. Tao, and Q. Du, "Spatial–spectral feature extraction via deep con-vLSTM neural networks for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 6, pp. 4237–4250, June 2020.

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

[131] X. Zhang, Y. Wang, N. Zhang et al., "Spectral-spatial three-dimensional convolutional neural network for hyperspectral image classification," *IEEE Access*, vol. 8, pp. 127167–127180, 2020.

[132] L. Mou, P. Ghamisi, and X. X. Zhu, "Deep recurrent neural networks for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 7, pp. 3639–3655, July 2017.

[133] "Types of RNN (recurrent neural network)," https://iq.opengenus.org/types-of-rnn/.

[134] "5 types of LSTM recurrent neural networks and what to do with them," https://www.exxactcorp.com/blog/Deep-Learning/5-types-of-lstm-recurrent-neural-networks-and-what-to-do-with-them.

[135] 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.

[136] S. Hao, W. Wang, and M. Salzmann, "Geometry-aware deep recurrent neural networks for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing, Early Access*, vol. 59, pp. 1–13, 2020.

[137] 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, June 2015.

[138] A. Mughees and L. Tao, "Multiple deep-belief-network-based spectral-spatial classification of hyperspectral images," *Tsinghua Science and Technology*, vol. 24, no. 2, pp. 183–194, April 2019.

[139] C. Chen, Y. Ma, and G. Ren, "Hyperspectral classification using deep Belief networks based on conjugate gradient update and pixel-centric spectral block features," *Ieee Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 13, pp. 4060–4069, 2020.

[140] 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, Feb. 2018.

[141] "The math behind GANs (generative adversarial networks)," https://towardsdatascience.com/the-math-behind-gans-generative-adversarial-networks-3828f3469d9c.

[142] "Introduction to generative adversarial networks (GANs): types, and applications, and implementation," https://heartbeat.fritz.ai/introduction-to-generative-adversarial-networks-gans-35ef44f21193.

[143] 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.

[144] H. Wang, C. Tao, J. Qi, H. Li, and Y. Tang, "Semi-supervised variational generative adversarial networks for hyperspectral image classification," in *Proceedings of the IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium*, IEEE, Yokohama, Japan, 28 July 2019.

[145] C. Tao, H. Wang, J. Qi, and H. Li, "Semisupervised variational generative adversarial networks for hyperspectral image classification," *Ieee Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 13, pp. 914–927, Feb. 2020.

[146] Z. Zhon, 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, July 2020.

[147] H. Liang, W. Bao, and X. Shen, "Adaptive weighting feature fusion approach based on generative adversarial network for hyperspectral image classification," *Remote Sensing*, vol. 13, p. 198, 2021.

[148] Z. Li, X. Zhu, Z. Xin, F. Guo, X. Cui, and L. Wang, "Variational generative adversarial network with crossed spatial and spectral interactions for hyperspectral image classification," *Remote Sensing*, vol. 13, p. 3131, 2021.

[149] M. Tsiakmaki, G. Kostopoulos, S. Kotsiantis, and O. Ragos, "Transfer learning from deep neural networks for predicting student performance," *Applied Sciences*, vol. 10, no. 6, p. 2145, 2020.

[150] J. Lin, R. Ward, and Z. J. Wang, "Deep transfer learning for Hyperspectral Image classification," in *Proceedings of the 2018 IEEE 20th International Workshop on Multimedia Signal Processing (MMSP)*, IEEE, Yokohama, Japan, 29 August 2018.

[151] 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, March 2019.

[152] X. He, Y. Chen, and P. Ghamisi, "Heterogeneous transfer learning for hyperspectral image classification based on convolutional neural network," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 5, pp. 3246–3263, May 2020.

[153] Y. Liu, L. Gao, C. Xiao, Y. Qu, K. Zheng, and A. Marinoni, "Hyperspectral image classification based on a shuffled group convolutional neural network with transfer learning," *Remote Sensing*, vol. 12, p. 1780, 2020.

[154] F. Xie, Q. Gao, C. Jin, and F. Zhao, "Hyperspectral image classification based on superpixel pooling convolutional neural network with transfer learning," *Remote Sensing*, vol. 13, p. 930, 2021.

[155] Y. Kumar, A. Koul, R. Singla, and M. F. Ijaz, "Artificial intelligence in disease diagnosis: a systematic literature review, synthesizing framework and future research agenda," *Journal of Ambient Intelligence and Humanized Computing*, vol. 13, pp. 1–28, 2022.

[156] "Data-driven cervical cancer prediction model with outlier detection and over-sampling methods. MF Ijaz, M Attique, Y Son," *Sensors*, vol. 20, no. 10, pp. 2809–76, 2020.

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

[158] J. Xia, J. Chanussot, P. Du, and X. He, "(Semi-) supervised probabilistic principal component analysis for hyperspectral remote sensing image classification," in *Proceedings of the 2012 4th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, IEEE, 4 June 2012.

[159] Y. Ren, L. Liao, S. J. Maybank, Y. Zhang, and X. Liu, "Hyperspectral image spectral-spatial feature extraction via tensor principal component analysis," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 9, pp. 1431–1435, 2017.

[160] S. Kutluk, K. Kayabol, and A. Akan, "Classification of hyperspectral images using mixture of probabilistic PCA models," in *Proceedings of the 2016 24th European Signal Processing Conference (EUSIPCO)*, IEEE, Budapest, Hungary, 29 August 2016.

[161] X. Kang, X. Xiang, S. Li, and J. A. Benediktsson, "PCA-based edge-preserving features for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 12, pp. 7140–7151, Dec. 2017.

[162] S. Chiang, C. Chang, and I. W. Ginsberg, "Unsupervised hyperspectral image analysis using independent component analysis," in *Proceedings of the IGARSS 2000. IEEE 2000 International Geoscience and Remote Sensing Symposium. Taking the Pulse of the Planet: The Role of Remote Sensing in Managing the Environment. Proceedings (Cat. No.00CH37120)*, IEEE, Honolulu, HI, USA, 24 July 2000.

[163] A. Villa, J. A. Benediktsson, J. Chanussot, and C. Jutten, "Independent component discriminant analysis for hyperspectral image classification," in *Proceedings of the 2010 2nd Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing*, IEEE, Reykjavik, Iceland, 14–16 June 2010.

[164] A. Villa, J. A. Benediktsson, J. Chanussot, and C. Jutten, "Hyperspectral image classification with independent component discriminant analysis," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 12, pp. 4865–4876, Dec. 2011.

[165] Y. Chen, L. Xu, Y. Fang et al., "Unsupervised bayesian subpixel mapping of hyperspectral imagery based on band-weighted discrete spectral mixture model and Markov random field," *IEEE Geoscience and Remote Sensing Letters*, vol. 18, no. 1, pp. 162–166, Jan. 2021.

[166] H. Chen, F. Miao, Y. Chen, Y. Xiong, and T. Chen, "A hyperspectral image classification method using multifeature vectors and optimized KELM," *Ieee Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 2781–2795, 2021.

[167] A. Saboori, H. Ghassemian, and F. Razzazi, "Active multiple kernel Fredholm learning for hyperspectral images classification," *IEEE Geoscience and Remote Sensing Letters*, vol. 18, no. 2, pp. 356–360, Feb. 2021.

[168] Y. Zhang, G. Cao, B. Wang, X. Li, P. Y. O. Amoako, and A. Shafique, "Dual sparse representation graph-based copropagation for semisupervised hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, 2021.

[169] Z. Lei, Y. Zeng, P. Liu, and X. Su, "Active deep learning for hyperspectral image classification with uncertainty

[170] K. Y. Ma and C. I. Chang, "Iterative training sampling coupled with active learning for semisupervised spectral-spatial hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, 2021.

[171] Q. Lu and L. Wei, "Multiscale superpixel-based active learning for hyperspectral image classification," *IEEE Geo-Science and Remote Sensing Letters*, vol. 19, 2021.

[172] R. Hang, Z. Li, Q. Liu, P. Ghamisi, and S. S. Bhattacharyya, "Hyperspectral image classification with attention-aided CNNs," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 3, pp. 2281–2293, March 2021.

[173] 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, 2021.

[174] R. Hang, F. Zhou, Q. Liu, and P. Ghamisi, "Classification of hyperspectral images via multitask generative adversarial networks," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 2, pp. 1424–1436, Feb. 2021.

[175] 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, March 2021.

[176] J. Wang, S. Guo, R. Huang, L. Li, X. Zhang, and L. Jiao, "dual-channel capsule generation adversarial network for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, 2021.

[177] 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, 2021.

[178] L. Fang, S. Li, X. Kang, and J. A. Benediktsson, "Spectral–spatial classification of hyperspectral images with a super-pixel-based discriminative sparse model," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 8, pp. 4186–4201, 2015.

[179] P. Ma, J. Ren, H. Zhao, G. Sun, P. Murray, and J. Zheng, "Multiscale 2-D singular spectrum analysis and principal component analysis for spatial–spectral noise-robust feature extraction and classification of hyperspectral images," *Ieee Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 1233–1245, 2021.

[180] D. M. S. Arsa, H. R. Sanabila, M. F. Rachmadi, A. Gamal, and W. Jatmiko, "Improving principal component analysis performance for reducing spectral dimension in hyperspectral image classification," in *Proceedings of the 2018 International Workshop on Big Data and Information Security (IWBIS)*, pp. 123–128, IEEE, Jakarta, Indonesia, 12 May 2018.

[181] M. Baisantry and A. K. Sao, "Band selection using segmented PCA and component loadings for hyperspectral image classification," in *Proceedings of the IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium*, pp. 3812–3815, IEEE, Yokohama, Japan, 28 July 2019.

[182] M. M. Hossain and M. A. Hossain, "Feature reduction and classification of hyperspectral image based on multiple kernel PCA and deep learning," in *Proceedings of the 2019 IEEE International Conference on Robotics, Automation, Artificial-intelligence and Internet-of-Things (RAAICON)*, pp. 141–144, IEEE, Dhaka, Bangladesh, 29 November 2019.

[183] D. Ruiz, B. Bacca, and E. Caicedo, "Hyperspectral images classification based on inception network and kernel PCA," *IEEE Latin America Transactions*, vol. 17, no. 12, pp. 1995–2004, December 2019.

[184] 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, Nov. 2018.

[185] J. Lin, L. Mou, X. X. Zhu, X. Ji, and Z. J. Wang, "Attention-aware pseudo-3-D convolutional neural network for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, 2021.

[186] V. Neagoe and P. Diaconescu, "CNN hyperspectral image classification using training sample augmentation with generative adversarial networks," . in *Proceedings of the 2020 13th International Conference on Communications (COMM)*, pp. 515–519, IEEE, Bucharest, Romania, 18 June 2020.

[187] X. Zhang, X. Jiang, J. Jiang, Y. Zhang, X. Liu, and Z. Cai, "Spectral-spatial and superpixelwise PCA for unsupervised feature extraction of hyperspectral imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, 2021.

[188] A. N. Abbasi and M. He, "Convolutional neural network with PCA and batch normalization for hyperspectral image classification," in *Proceedings of the IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium*, pp. 959–962, IEEE, Yokohama, Japan, 28 July 2019.

[189] M. R. Haque and S. Z. Mishu, "Spectral-spatial feature extraction using PCA and multi-scale deep convolutional neural network for hyperspectral image classification," in *Proceedings of the 2019 22nd International Conference on Computer and Information Technology (ICCIT)*, pp. 1–6, IEEE, Dhaka, Bangladesh, 18 December 2019.

[190] W. Sun, G. Yang, J. Peng, and Q. Du, "Lateral-slice sparse tensor robust principal component analysis for hyperspectral image classification," *IEEE Geoscience and Remote Sensing Letters*, vol. 17, no. 1, pp. 107–111, Jan. 2020.

[191] A. N. Abbasi and M. He, "CNN with ICA-PCA-DCT joint preprocessing for hyperspectral image classification," in *Proceedings of the 2019 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, pp. 595–600, IEEE, Lanzhou, China, 18 November 2019.

[192] M. Baisantry, A. K. Sao, and D. P. Shukla, "Band selection using combined divergence–correlation index and sparse loadings representation for hyperspectral image classification," *Ieee Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 13, pp. 5011–5026, 2020.

[193] S. Jia, J. Hu, Y. Xie, L. Shen, X. Jia, and Q. Li, "Gabor cube selection based multitask joint sparse representation for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 6, pp. 3174–3187, June 2016.

[194] Y. Zhan, J. Qin, T. Huang et al., "Hyperspectral image classification based on generative adversarial networks with feature fusing and dynamic neighborhood voting mechanism," in *Proceedings of the IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Sympo-sium*, pp. 811–814, IEEE, Yokohama, Japan, 28 July 2019.

[195] A. A. Joy and M. A. M. Hasan, "A hybrid approach of feature selection and feature extraction for hyperspectral image classification," in *Proceedings of the 2019 International Conference on Computer, Communication, Chemical,* *Materials and Electronic Engineering (IC4ME2)*, pp. 1–4, IEEE, Rajshahi, Bangladesh, 11 July 2019.

[196] U. A. M. E. Ali, M. A. Hossain, and M. R. Islam, "Analysis of PCA based feature extraction methods for classification of hyperspectral image," in *Proceedings of the 2019 2nd International Conference on Innovation in Engineering and Technology (ICIET)*, pp. 1–6, Dhaka, Bangladesh, 23 December 2019.

[197] Y. Zhan, K. Wu, W. Liu et al., "Semi-supervised classification of hyperspectral data based on generative adversarial networks and neighborhood majority voting," in *Proceedings of the IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium*, pp. 5756–5759, IEEE, Valencia, Spain, 22 July 2018.

[198] A. I. Champa, M. F. Rabbi, and N. Banik, "Improvement in hyperspectral image classification by using hybrid subspace detection technique," in *Proceedings of the 2019 International Conference on Sustainable Technologies for Industry 4.0 (STI)*, pp. 1–5, IEEE, Dhaka, Bangladesh, 24 December 2019.

[199] H. Fu, G. Sun, J. Ren, A. Zhang, and X. Jia, "Fusion of PCA and segmented-PCA domain multiscale 2-D-SSA for effective spectral-spatial feature extraction and data classification in hyperspectral imagery," *IEEE Transactions on Geoscience and Remote Sensing*.

[200] G. Y. Chen, "Multiscale filter-based hyperspectral image classification with PCA and SVM," *Journal of Electrical Engineering*, vol. 72, no. 1, pp. 40–45, 2021.

[201] 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, March 2018.

[202] K. Bhardwaj, A. Das, and S. Patra, "Spectral-spatial active learning with superpixel profile for classification of hyperspectral images," in *Proceedings of the 2020 6th International Conference on Signal Processing and Communication (ICSC)*, pp. 149–155, IEEE, Noida, India, 5 March 2020.

[203] 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, July 2020.

[204] M. E. Paoletti, J. M. Haut, J. Plaza, and A. Plaza, "Training capsnets via active learning for hyperspectral image classification," in *Proceedings of the IGARSS 2020 - 2020 IEEE International Geoscience and Remote Sensing Symposium*, pp. 40–43, IEEE, Waikoloa, HI, USA, 26 September 2020.

[205] 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, Nov. 2018.

[206] X. Wang, K. Tan, and Y. Chen, "CapsNet and triple-GANs towards hyperspectral classification," in *Proceedings of the 2018 Fifth International Workshop on Earth Observation and Remote Sensing Applications (EORSA)*, pp. 1–4, IEEE, Xi'an, China, 18 June 2018.

[207] 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, Sept. 2019.

[208] O. Okwuashi and C. E. Ndehedehe, "Deep support vector machine for hyperspectral image classification," *Pattern Recognition*, vol. 103, Article ID 107298, 2020.

[209] F. I. Alam, J. Zhou, A. W. Liew, X. Jia, J. Chanussot, and Y. Gao, "Conditional random field and deep feature learning

for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 3, pp. 1612–1628, March 2019.

[210] Y. Cao, J. Mei, W. Yuebin et al., "SLCRF: subspace learning with conditional random field for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, 2021.

[211] Y. Liang, X. Zhao, A. J. X. Guo, and F. Zhu, "Hyperspectral image classification with deep metric learning and conditional random field," *IEEE Geoscience and Remote Sensing Letters*, vol. 17, no. 6, pp. 1042–1046, June 2020.

[212] Y. Wang, J. Mei, L. Zhang et al., "Self-supervised feature learning with CRF embedding for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 5, pp. 2628–2642, May 2019.

[213] V. Andrejchenko, W. Liao, W. Philips, and P. Scheunders, "Decision fusion framework for hyperspectral image classification based on Markov and conditional random fields," *Remote Sensing*, vol. 11, no. 6, p. 624, 2019.

[214] J. V. Rissati, P. C. Molina, and C. S. Anjos, "Hyperspectral image classification using random forest and deep learning algorithms," in *Proceedings of the 2020 IEEE Latin American GRSS & ISPRS Remote Sensing Conference (LAGIRS)*, p. 132, IEEE, Santiago, Chile, 22 March 2020.

[215] Y. Zhang, G. Cao, X. Li, and B. Wang, "Cascaded random forest for hyperspectral image classification," *Ieee Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 11, no. 4, pp. 1082–1094, April 2018.

[216] 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.

[217] T. Li, J. Leng, and L. D. C. N. R. Kong, "Deep cube CNN with random forest for hyperspectral image classification," *Multimedia Tools and Applications*, vol. 78, pp. 3411–3433, 2019.

[218] A. Wang, Y. Wang, and Y. Chen, "Hyperspectral image classification based on convolutional neural network and random forest," *Remote Sensing Letters*, vol. 10, no. 11, pp. 1086–1094.

[219] 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, 2021.

[220] Y. Ding, Y. Guo, Y. Chong, S. Pan, and J. Feng, "Global consistent graph convolutional network for hyperspectral image classification," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–16, Article ID 5501516, 2021.

[221] X. He, Y. Chen, and P. Ghamisi, "Dual graph convolutional network for hyperspectral image classification with limited training samples," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, 2021.

[222] J. Bai, B. Ding, Z. Xiao, L. Jiao, H. Chen, and A. C. Regan, "Hyperspectral image classification based on deep attention graph convolutional network," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, 2021.

[223] Q. Liu, L. Xiao, J. Yang, and Z. Wei, "CNN-enhanced graph convolutional network with pixel- and superpixel-level feature fusion for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, 2021.

[224] A. Signoroni, M. Savardi, A. Baronio, and S. Benini, "Deep learning meets hyperspectral image analysis: a multidisciplinary review," *Journal of Imaging*, vol. 5, no. 5, p. 52, 2019.