



# An extensive review of hyperspectral image classification and prediction: techniques and challenges

Ganji Tejasree<sup>1</sup> · Loganathan Agilandeewari<sup>1</sup>

Received: 10 July 2023 / Revised: 9 November 2023 / Accepted: 29 January 2024

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2024

## Abstract

Hyperspectral Image Processing (HSIP) is an essential technique in remote sensing. Currently, extensive research is carried out in hyperspectral image processing, involving many applications, including land cover classification, anomaly detection, plant classification, etc., Hyperspectral image processing is a powerful tool that enables us to capture and analyze an object's spectral information with greater accuracy and precision. Hyperspectral images are made up of hundreds of spectral bands, capturing an immense amount of information about the earth's surface. Accurately classifying and predicting land cover in these images is critical to understanding our planet's ecosystem and the impact of human activities on it. With the advent of deep learning techniques, the process of analyzing hyperspectral images has become more efficient and accurate than ever before. These techniques enable us to categorize land cover and predict Land Use/Land Cover (LULC) with exceptional precision, providing valuable insights into the state of our planet's environment. Image classification is difficult in hyperspectral image processing because of the large number of data samples but with a limited label. By selecting the appropriate bands from the image, we can get the finest classification results and predicted values. To our knowledge, the previous review papers concentrated only on the classification method. Here, we have presented an extensive review of various components of hyperspectral image processing, hyperspectral image analysis, pre-processing of an image, feature extraction and feature selection methods to select the number of features (bands), classification methods, and prediction methods. In addition, we also elaborated on the datasets used for classification, evaluation metrics used, various issues, and challenges. Thus, this review article will benefit new researchers in the hyperspectral image classification domain.

**Keywords** Hyperspectral image processing · Spectral bands · Hyperspectral image classification · Land cover classification · LULC change prediction

---

✉ Loganathan Agilandeewari  
agila.l@vit.ac.in

Ganji Tejasree  
ganji.tejasree2020@vitstudent.ac.in

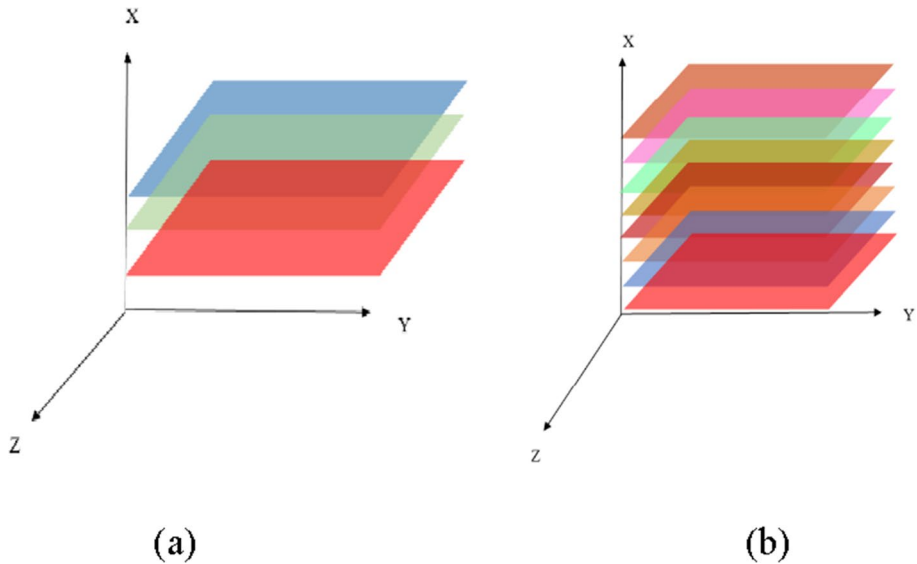
<sup>1</sup> School of Computer Science Engineering and Information Systems, Vellore Institute of Technology, Vellore-632014, Tamil Nadu, India

## 1 Introduction

Hyperspectral Imaging (HSI) development traces back to the early 1970s when NASA started developing airborne imaging spectrometers to study the Earth's surface. The first commercial HSI systems became available in the early 1990s, and the technology has since become more affordable and accessible. Hyperspectral imaging (HSI) captures the electromagnetic spectrum in detail, providing information on the physical and chemical properties of objects. It's non-invasive and useful in agriculture, mineralogy, and environmental monitoring. HSI's spectral signature identification helps understand underlying phenomena. The data collected by HSI can be used to identify and distinguish between different materials, detect changes in the environment, monitor the health of crops, and much more. It measures the reflected light from an object or scene and is also known as imaging spectroscopy [1]. Hyperspectral imaging (HSI) is an advanced technology capable of extracting valuable information from images. HSI sensors capture a large number of spectral bands, allowing for a much more detailed analysis compared to traditional RGB imaging. This means that HSI sensors can provide a greater level of detail in the scene being analyzed. This nuanced analysis offers a more comprehensive understanding of the image, allowing for better-informed decision-making [2].

Hyperspectral Imaging (HSI) is the most trending approach, primarily used for analyzing the earth using Remote Sensing (RS). Remote sensing gathers data on objects without physical contact, using technology to measure properties like temperature and radiation. It's valuable in fields like environmental monitoring, geology, and urban planning. It allows the collection of spectral, geographical, and temporal data about physical objects, regions, or areas under inquiry; it has many applications in several aspects of earth science, like agriculture, geology, and environmental monitoring [3]. RGB images, consisting of three dimensions representing color information, were commonly used before hyperspectral imaging (HSI). However, HSI captures spectral information across a range of electromagnetic wavelengths, providing a more detailed spectral profile of the object being imaged. HSI images are captured using multiple narrow spectral bands, which are then combined to create a three-dimensional data cube that contains information on the spectral profile of each pixel. HSI imaging is a valuable tool for a wide range of applications, from remote sensing to medical imaging. These color values combine RGB intensities displayed on a color plane. After RGB images, Multispectral Images (MSI) came into the world, they captured more spectral bands than RGB images. Multispectral sensors usually capture illumination energy contemplated from the objects over the earth's surface. It typically has 3 to 10 different spectral bands. Examples of bands in these sensors include visible green, red, blue, invisible infrared, etc. [4]. Figure 1 represents the RGB Image and Hyperspectral images.

Remote sensing data for hyperspectral imaging will be collected using platforms such as aircraft satellites, balloons, rockets, space shuttles, etc.... Spaceborne and airborne sensors are most used for capturing hyperspectral images. In spaceborne images, the sensors will capture the images from the space station. Airborne images will be captured through aircraft [6]. The Airborne sensors are an Airborne Visible/Infrared Imaging Spectrometer (AVIRIS), Compact Airborne Spectrographic Imager (CASI), Hyperspectral Digital Image Collection Experiment (HYDIC), Digital Airborne Imaging Spectrometer (DAIS), Push-Broom Hyperspectral Imager (PHI), HyMap, Airborne Prism Experiment (APEX), Modular Airborne IMAGING Spectrometer (MAIS), and UAS (Unmanned Aircraft Systems) [7]. The spaceborne sensors are MERIS (Medium Resolution Imaging Spectrometer),


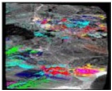

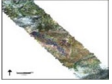


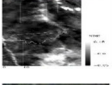


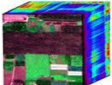


**Fig. 1** **a** RGB Images, **b** Hyperspectral Images [5]

HYSI(INDIA), MODIS (Moderate Resolution Imaging Spectrometer), Hyperion, Hyperspectral Imager, NEMO (Naval Earth Map Observer), and OrbiView-4 [8]. Details about sensors are given in Table 1.

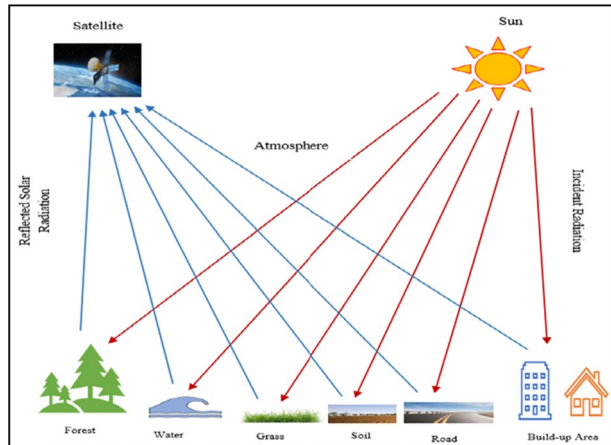
The data collected from the airborne and spaceborne sensors will be reflected in spectral bands, and those spectral bands will form a hyperspectral data cube. Hyperspectral sensors will capture images and process an image towards an exceedingly massive number of wavelengths. HSI has become an effective method to observe the earth, which can contribute to the total spectral specifications of an object in supplement to restricted bands ranging from  $0.4 \mu\text{m}$  to  $10 \mu\text{m}$ . It does not quickly provide the target's position information [9]. Every hyperspectral image has its own spatial, spectral, and temporal resolution. The Ground Sampling Distance (GSD) is a measure of the spatial resolution of an image, which determines the ability to differentiate between adjacent objects with high accuracy. GSD is the smallest detectable object size in ground-based hyperspectral imaging, typically ranging from 1 to 10 m [10]. Spectral resolution refers to an image's capacity to differentiate between various wavelengths of light, measured in nanometres (nm). Hyperspectral images are a type of remote sensing imagery that contains a vast amount of detailed spectral information. With their ability to capture hundreds or thousands of narrow and contiguous spectral bands, they provide an in-depth understanding of the target object's surface characteristics, chemical composition, and physical properties. This enables the identification of a vast array of materials and features that would be impossible to detect with conventional RGB imaging [11]. Temporal resolution refers to the frequency at which images are captured, measured in seconds, minutes, or hours. Based on the intended application, hyperspectral images can be acquired with various temporal resolutions. For example, hyperspectral images of crops may be captured every few days to monitor crop health and development. On the other hand, hyperspectral images of urban areas may be captured less frequently, such as once a month or once a year [12]. Figure 2 portrays the electromagnetic radiation that is either reflected or emitted by the surface of the Earth.

**Table 1** Explanation of hyperspectral sensors

S. No	Reference	Sensor	Introduced year	Number of Spectral bands	Operating wavelength range (nm)	Image
1	[16]	Reflecting Optics System Imaging Spectrometer (ROSIS)	1992	115	430-860	
2	[17]	Airborne Visible/ Infrared Imaging Spectrometer (AVIRIS)	Mid-1990	224	400-2500	
3	[18]	Hyperspectral Imaging Camera (HySI)	2018	32	400-950	
4	[19]	Compact Airborne Spectro-graphic Imager (CASI)	1988	288	400-900	
5	[20]	Hyperspectral Digital Imagery Collection Experiment (HYDICE)	1994	210	400-2500	
6	[21]	HyMap	-	126	400-2500	
7	[22]	Moderate Resolution Imaging Spectra radiometer (MODIS)	1999	36	400-1440	
8	[23]	Hyperion	1998	220	357-2576	
9	[24]	Compact High-Resolution Imaging Spectrometer (CHRIS)	2001	19	400-1050	
10	[25]	Headwall II Nano-Hyperspec	2014	270	400-1000	

In hyperspectral imaging, a band refers to a collection of wavelengths that are detected by the hyperspectral sensor. Unlike traditional RGB imaging, hyperspectral sensors can capture hundreds or even thousands of bands, providing a highly detailed and nuanced scene analysis [13]. A hyperspectral image consists of several bands, each containing information about a specific range of wavelengths. For instance, one band may capture information about the red wavelength range, while another may capture information about the green wavelength range. For example, Visible bands are sensitive to the colors that we see with our eyes. Near-infrared (NIR) bands are sensitive to the amount of reflected sunlight, ranging from 400–700 nm (nm). Shortwave infrared (SWIR) bands are sensitive to the water content of

**Fig. 2** The source of remote sensing image capturing is taken by [15]

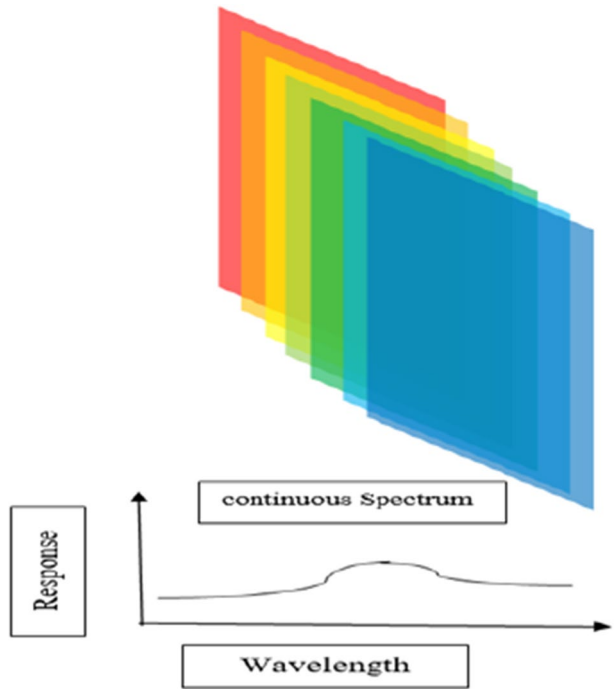


materials, ranging from 700–1300 nm. Midwave infrared (MWIR) bands are sensitive to the temperature of materials, ranging from 1300–2500 nm. Longwave infrared (LWIR) bands are sensitive to the heat emitted by objects, ranging from 2500–5000 nm. By analyzing the information obtained through hyperspectral imaging, researchers can gain insights into the composition, properties, and conditions of materials in the scene [14].

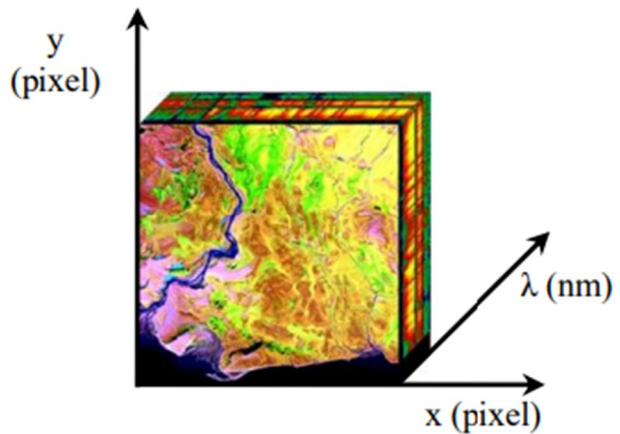
Hyperspectral imaging creates spectral bands by dividing the light from the scene into its constituent wavelengths. This can be accomplished using prisms, which refract light based on wavelength, or diffraction gratings, which diffract light based on wavelength. Whispering Gallery Mode (WGM) resonators are microcavities that capture and circulate light multiple times. The resonant wavelength of a WGM resonator depends on its size and shape, which can be used to create a filter that permits only light of a specific wavelength to pass through. Once the light is dispersed into its wavelengths, a detector array captures it. The detector array consists of numerous individual detectors, each sensitive to a different wavelength of light. The detector array records the light's intensity at each wavelength, creating a hyperspectral image cube [26]. The representation of spectral bands are shown in Fig. 3.

A hyperspectral data cube is a 3D dataset that captures the spectral reflectance of every pixel in a scene. It is created by a hyperspectral sensor, which captures scene images at numerous wavelengths. Hyperspectral sensors can be mounted on aircraft, satellites, or even handheld devices. Using a prism or diffraction grating, light is divided into wavelengths. Each sensor element then detects a different wavelength of light. After capturing multiple images of a scene, the process of stacking these images together creates a 3D data cube. Every pixel in this cube holds a unique representation of a moment captured in time. The  $x$  and  $y$  dimensions hold the spatial coordinates, while the  $z$  dimension encapsulates the essence of the captured light wavelength. In other words, the cube provides a comprehensive view of the scene in three dimensions, combining both spatial and spectral information. The hyperspectral datacube is shown in Fig. 4. Hyperspectral data cubes are large and intricate datasets that provide abundant information about the scene. The data gathered can be employed to accurately detect and visually represent diverse substances such as minerals, plant life, and bodies of water. Moreover, hyperspectral data cubes are employed to keep track of environmental changes like deforestation and pollution [28]. The following are the steps involved in creating a hyperspectral data cube:

**Fig. 3** The representation of spectral bands [27]



**Fig. 4** Hyperspectral data cube [29]



- Step 1: Collect images of the scene at numerous wavelengths using a hyperspectral sensor.
- Step 2: Calibrate the images to eliminate any sensor noise or artifacts.
- Step 3: Geo-reference the images to associate them with a known coordinate system.
- Step 4: Stack the images together to create a 3D data cube.
- Step 5: Process the data cube to remove atmospheric effects and enhance the signal-to-noise ratio.

In the field of hyperspectral imaging, a spectral signature represents a definitive and unmistakable pattern of light that an object or material emits or reflects at varying wavelengths. It works like a fingerprint, enabling materials to be identified and classified in hyperspectral images. Hyperspectral sensors excel in capturing hundreds of spectral bands, enabling a highly detailed and nuanced analysis compared to traditional RGB imaging. By scrutinizing the spectral signatures of various pixels in a hyperspectral image, scientists can gain insight into the materials present in the scene, including their composition, structure, and temperature. Several factors influence spectral signatures, including the material's composition, structure, surface texture, and illumination conditions. For instance, different vegetation types have distinct spectral signatures due to their varying pigments and leaf structures. Similarly, minerals of different types exhibit different spectral signatures owing to their disparate chemical compositions [30] (Fig. 5).

Hyperspectral Imaging (HSI) technology is an advanced imaging technique that comprises hundreds of spectral bands, which can be confidently utilized across various fields. These fields include environmental monitoring, military surveillance, urban planning, precision agriculture, seed viability studies, pharmaceuticals, biotechnology, oil and gas, medical diagnosis, thin films, and forensic science [31]. Detailed information about each application will be explained in section 3 We can use hyperspectral images for classification and prediction. In classification, we can classify the land cover and so on. In the medical field, we can classify various types of tumors, blood cells, etc. By using prediction methods, we can predict deforestation, etc. The characteristics of RGB, multispectral, and hyperspectral imaging are listed in Table 2.

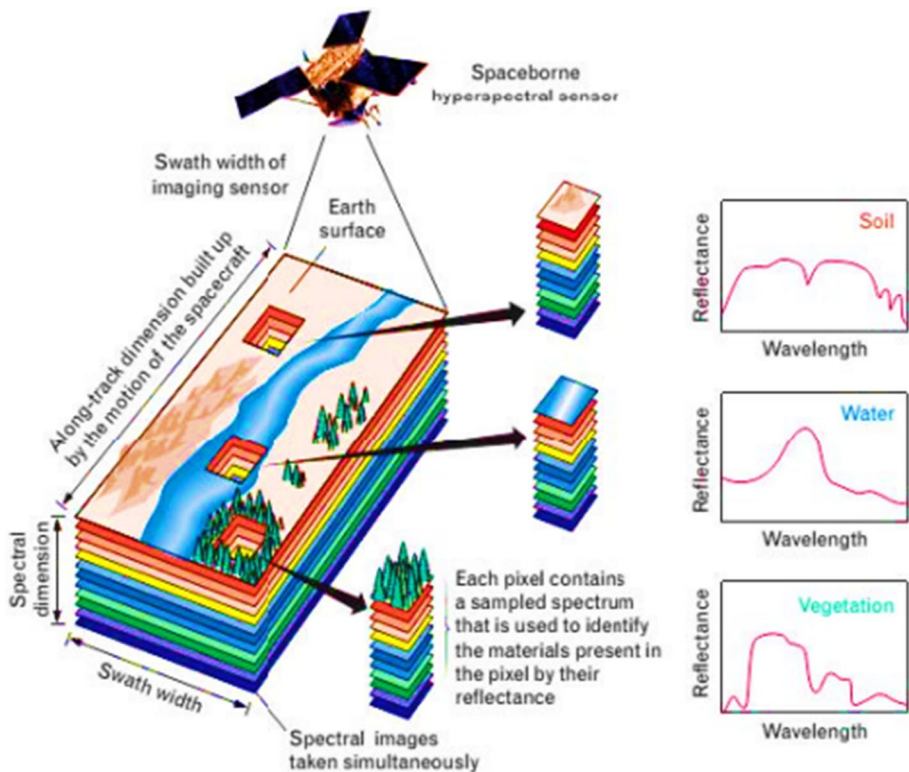


Fig. 5 Spectral signatures of soil, vegetation, and water [30]



**Table 2** Characteristics of RGB, multispectral, and hyperspectral imaging

Characteristics	RGB	Multispectral	Hyperspectral
Number of spectral bands	3	4–10	Hundreds to thousands
Spectral Resolution	Low	Medium	High
Spatial Resolution	Typically high	Typically high	Variable
Applications	Photography, videography	Remote sensing, environmental monitoring, agricultural	Research, commercial applications in a wide range

Hyperspectral imaging is a highly advanced technology that uses a range of wavelengths of light to capture detailed information about the composition of a land surface. It can identify and classify various types of vegetation, soil, water bodies, and man-made structures with great accuracy. This makes it a valuable tool for analyzing and monitoring land use and land cover changes over time, which can aid in making informed decisions related to environmental management and resource planning. It captures detailed images by using various light wavelengths and allows for informed decision-making in various applications like urban planning, crop monitoring, and natural resource management. Hyperspectral imaging offers detailed information on the spectral signatures of diverse materials, which can be effectively leveraged to develop machine learning algorithms for the accurate classification of various land cover types. Within hyperspectral imaging, there exist several methods for LULC classification. One of the most frequently used techniques is the utilization of a spectral library, which represents a database containing the spectral signatures of diverse materials. To classify pixels into different land cover types, a plethora of machine learning and deep learning algorithms are employed. These algorithms compare the spectral signatures of the pixels with those present in the spectral library [32]. Another standard method is using machine learning algorithms trained using labeled training data. Labeled training data is a set of pixels manually classified into different land cover types. Once the LULC classification model is trained, it can classify the pixels in new hyperspectral images. The LULC classification map generated as a result can be effectively employed to identify and map different types of land cover present within the scene. LULC classification in hyperspectral imaging is an advancing field with new applications being constantly discovered. It has several applications in today's world, such as identifying areas of fields that are stressed or diseased, tracking the spread of invasive algae blooms in lakes and rivers, mapping urban areas, identifying areas for development, and mapping forests, wetlands, and other natural resources [33].

The significance of Hyperspectral Imaging (HSI) is it is a valuable tool for extracting useful information from images. However, HSI has several challenges that need to be addressed, it produces a large amount of data, making it challenging to store, process, and analyze them. The spectral signatures of materials can vary depending on factors like illumination, atmospheric conditions, and surface roughness. This can make it difficult to identify and classify materials in HSI images accurately. Labeled data is crucial for training machine learning models to identify and classify materials in HSI images. Unfortunately, acquiring labeled HSI data can be difficult and expensive. Many of the algorithms used to process and analyze HSI images are computationally intensive. This makes it challenging to use HSI in real-time applications [34].

In recent years, numerous review articles have been published regarding the classification of Hyperspectral Images. Articles such as, [35] present a comprehensive and confident analysis of machine-dependent technologies and deep learning methods that are most effective in hyperspectral image classification. A thorough review of the literature provides valuable insights



and conclusions that are essential for researchers to understand the intricate relationship between machine learning and hyperspectral imaging. The authors in [36], explore the potential of combining hyperspectral imaging with deep learning to solve tasks across various application fields while highlighting potentialities and critical issues related to the development trends. In [3], this review offers a comprehensive and confident analysis of deep learning techniques used for hyperspectral image classification. It provides an in-depth evaluation of the strengths and weaknesses of the most widely used classifiers and presents reliable quantitative results for easy comparison. As a result, this review presents a definitive understanding of deep learning techniques for hyperspectral image classification. And also, the authors in [37], the article delves into the application of Convolutional Neural Networks (CNNs) for the classification of Hyperspectral Images (HSIs). A comprehensive analysis is conducted, comparing the efficacy of four distinct CNN models in capturing both spectral and spatial features, as well as their combination. Furthermore, the article provides insightful recommendations for future enhancements that could elevate the performance of these models. After conducting our research, we have found that numerous review articles exist regarding the classification of Hyperspectral Images (HSI) in a broad sense. Regrettably, none of these articles have delved into a comprehensive analysis of the complete HSI processing workflow. This workflow encompasses Hyperspectral Image acquisition, Hyperspectral data cube generation, Pre-processing, Feature Extraction, Band Selection, Classification, Prediction, benchmark datasets, and quality metrics. We aim to address this gap in research by providing a single, detailed article covering all aspects of the HSI processing workflow. Here, we are giving the complete details about the various sensors used for hyperspectral image acquisition, the methods involved in the pre-processing of the acquired hyperspectral image, feature extraction methods, band selection methods, classification methods, prediction methods, various benchmark datasets with their complete details, and quality metrics are elaborated. By reading our review paper, researchers can understand all the steps involved from image acquisition to prediction with quality metrics.

## 1.1 Inclusion criteria for paper selection

The process of reviewing studies involves selecting and pre-processing literature that meets the defined selection criteria. This initial step is crucial in ensuring that the review is based on relevant and appropriate studies. To ensure the best possible outcome, a thorough and methodical approach was employed, involving a series of steps to carefully evaluate and determine the suitability of the literature at hand.

Step 1: To begin our research, we conducted a thorough search of various digital portals, including ACM, IEEE, Springer, Elsevier, and Wiley, to identify the most appropriate journals and conferences for our study. We examined numerous studies published in different journals and conference proceedings to ensure we met our selection criteria. We searched for studies in several prominent journals such as IEEE Transactions on Geoscience and Remote Sensing, IEEE Geoscience and Remote Sensing Letters, IEEE Transactions on Image Processing, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, Computational Intelligence and Neuroscience, Remote Sensing Letters, Remote Sensing, Sensors, International Journal of Remote Sensing, Remote Sensing of Environment, arXiv preprint arXiv, International Journal of Applied Earth Observation and Geoinformation, Communications & Signal Processing, Ieee Access, among others. Our thorough search ensured we had access to a broad range of relevant studies and publications for our research.

Step 2: To carry out a comprehensive review of hyperspectral imaging, we conducted a thorough search for relevant studies spanning from 2001 to February 2023. We utilized a range of keywords including hyperspectral imaging, remote sensing, hyperspectral image classification, hyperspectral image prediction, hyperspectral imaging applications, hyperspectral image land use and cover classification, as well as hyperspectral image feature extraction and band selection. We then examined the abstracts of the identified papers to extract concise and relevant clusters while disregarding any irrelevant studies.

Step 3: During this step, we assessed a group of studies that were chosen in the previous step. We conducted a thorough analysis of the studies and used clustering analysis to narrow down the selection. After examining the introduction and conclusion of the studies that were selected in the previous step, we decided which ones to include in this review. Finally, we excluded any studies that did not meet our criteria.

## 1.2 Exclusion criteria for paper elimination

Conducting a systematic literature review is a crucial step in gaining a thorough understanding of the current trends, challenges, and future research areas in a particular field. In our review of hyperspectral imaging, we established specific criteria to ensure that only primary research papers that met our standards were included. We categorized the papers based on their datasets, years of publication, application areas, deep learning techniques, and features used. Our search for relevant studies was comprehensive, and we looked through various journals and conference proceedings from digital portals like ACM, IEEE, Springer, Elsevier, Wiley, and others. We excluded studies that were not written in English, did not have the domain in their title, or did not meet our specific criteria. Our primary focus was to identify how hyperspectral imaging techniques are being utilized in different domains. Our comprehensive analysis of the available literature on hyperspectral imaging has yielded insightful findings regarding the current state of research and the trajectory it is taking.

Our review article's contribution is providing a fundamental understanding of hyperspectral imaging for researchers. This paper thoroughly discusses the specific approaches employed at each level, from the image capture stage to the accuracy validation step of the hyperspectral image classification process. The essential structure and potential framework for hyperspectral image classification are explained briefly. In addition, the paper sheds light on some notable challenges and promising directions for the field. It underscores the need for developing sophisticated techniques capable of addressing the high complexity and dimensionality of hyperspectral data, while also highlighting the importance of creating robust methods that can effectively handle noise and other artifacts present in hyperspectral images.

This review paper is structured in a manner that allows for clear and concise presentation of information. The organization of the paper will be explained in the following sections, providing readers with a comprehensive understanding of the content to be covered: Section 2 gives the related work from existing hyperspectral image classification and prediction models. Section 3 elaborates on the various applications of hyperspectral imaging. In Section 4, we explain the different pre-processing techniques that are utilized for satellite and medical images. Section 5 elaborates on the feature extraction algorithms. Section 6 expands the feature or band selection methods. Section 7 illustrates hyperspectral image classification approaches. Section 8 demonstrates hyperspectral image prediction techniques. Section 9 picturizes the benchmark databases used. Section 10 qualifies the metrics used for calculating the classification accuracy. Section 11 gives the open issues

and challenges. Section 12 discusses the hyperspectral image classification existing methods and their performance. Section 13 concludes the article.

## 2 Related works

This section reviews existing Hyperspectral Image Classification (HSIC) and Prediction papers. We have divided the literature into the following: classification and prediction; in classification again, we divided Traditional Machine Learning and Neural Networks; in Traditional Machine Learning, we divided into supervised, unsupervised, and semi-supervised classification methods. In neural networks, we again divided into Traditional Neural Networks and Deep Learning. And then into a Convolutional neural network. Table 3 provides a comprehensive summary of various supervised classification techniques that have been used for Hyperspectral image classification, Table 4 provides a literature review of existing methods based on unsupervised classification techniques for Hyperspectral image classification, Table 5 offers a comprehensive literature review of semi-supervised classification techniques. Additionally, Table 6 provides a thorough examination of attention-based classification techniques for the same. Moving forward, Table 7 delves into methods based on CNN classification techniques, while Table 8 presents a literature review of prediction techniques for Hyperspectral image classification.

In [38], the article provides a detailed overview of the current pansharpening techniques used in the fusion of hyperspectral and panchromatic images. It examines the various categories of pansharpening techniques and offers an evaluation of the advantages and limitations of existing methods. A highly effective technique for merging high spatial resolution images with low-resolution images, using a state-of-the-art Fuzzy and Gyration Transform (GT) based image fusion method is proposed in [39]. It uses a Genetic Algorithm to tune the required fuzzification parameters and maximizes the overall entropy. Quantitative analysis shows that the proposed method has better structural detail, spatial resolution, and spectral information. A model for restoring visibility in hazy remote sensing images is proposed, which is constantly being improved and developed [40]. The proposed image restoration model is built on a fusion-based transmission map, a hybrid constraint-based variational model, and a dynamic differential evolution to optimize the control parameters. Through rigorous testing on 50 synthetic benchmarks and 50 real-life remote sensing images, the model has demonstrated superior performance when compared to other existing restoration models.

A robust and efficient classification methodology that seamlessly integrates Principal Component Analysis (PCA), Local Binary Pattern (LBP), and Back Propagation Neural Network (BPNN) is presented in [41]. This cutting-edge approach ensures superior accuracy and reliability in classification tasks. It is tested on three publicly available hyperspectral datasets and achieves satisfactory accuracy. The classification of hyperspectral images is often hindered by a lack of labeled data. To address this challenge, researchers have recently introduced a deep hybrid multi-graph neural network (DHMG) in [42]. This novel approach employs two distinct graph filters, a dense network, and a GraphSAGE-based network to refine the graph features. Extensive experimentation has demonstrated that the DHMG model outperforms current state-of-the-art models. The proposed method leverages the wavelet transform to extract both spatial and spectral information, resulting in an effective solution for hyperspectral image classification. The extracted information is then fused and used to classify the images via a Support Vector Machine (SVM) classifier [43]. Experiments show that this method is effective compared to conventional approaches. In [44], the authors present

**Table 3** Literature review of existing methods based on supervised classification techniques for Hyperspectral image classification

S. No	Ref	Method	Dataset	Metrics (OA)	Remarks
1	[55]	SVM	Indian Pines, Salinas, Pavia University	99.73	By performing simple operations on HSI data cubes, the proposed method unlocks new levels of classification accuracy and speed
2	[56]	SVM	Realtime	95.24	One possible way to enhance the accuracy of the proposed 3D-CNN model could be to consider incorporating both spectral and spatial information from the HSI data
3	[57]	SVM	Indian Pines, Salinas, Pavia University	96.74	This approach's main strength is its ability to incorporate spectral and spatial characteristics, significantly improving classification accuracy
4	[58]	SVM	Indian Pines, Salinas, KSC	93.14	This proposed method analysis can be utilized to combine various models to improve their classification performance
5	[59]	SVM	92AV3C		To mitigate the risk of over-fitting caused by multiple adjustable parameters, it is imperative to test these algorithms on a broader range of labeled multiclass datasets. This will enable us to establish a robust understanding of the algorithm's performance and its generalizability, thereby contributing significantly to the advancement of the field
6	[60]	SVM-based Nested Sliding Window (NSW)	Indian Pines, Salinas, Pavia University	91.40	When the training set samples are abundant, obtaining advantages in classification accuracy with NSW-PCA-SVM becomes difficult. Mainly, the NSW method demands more computational time

Table 3 (continued)

S. No	Ref	Method	Dataset	Metrics (OA)	Remarks
7	[61]	SVM	Real-time data	93.17	The proposed method effectively identifies damaged tree crowns, making it a valuable tool for monitoring forest health and assessing pests and diseases on a larger scale
8	[62]	Tree based classifier	Indian Pines	97.93	Since PCA only works on linear datasets, it may not achieve better accuracy. To overcome this, Kernel PCA (KPCA) can be used
9	[63]	Feature fusion and multi-layered gradient boosting decision tree model (FF-DT)	Pavia university, Indian pines, Salinas	99.80	The proposed model demonstrates exceptional performance by effectively utilizing the strengths of tree-based models and EMP features to extract spatial and spectral information from hyperspectral images. In comparison to other classical models, the proposed model outperforms them in various aspects, which highlights its potential to serve as an effective solution
10	[64]	Gradient boosting decision tree regression (GBDTR)	Xunsi River	97	Besides the proposed algorithm, the researchers also investigate the potential of deep learning techniques and UAV-borne hyperspectral imagery for water remote sensing
11	[65]	Decision Tree	Real-time data	86	It is concluded that even the minimum detectable Nd concentration in a carbonate sample is significant
12	[66]	Random forest	GF-5 and EO-1(real-time data)	Efficient results	To be more precise, the FSSRF technique employs Principal Component Analysis (PCA) to compress the hyperspectral bands of temporally adjacent HSIs

Table 3 (continued)

S. No	Ref	Method	Dataset	Metrics (OA)	Remarks
13	[67]	Exponentially Weighted Random Forest (EWRf)	Indian pines, Pavia university	95.01	The algorithm proposed in this study was evaluated against established methods on two distinct hyperspectral datasets. The results of our experiments demonstrate that the proposed algorithm is a competitive alternative for hyperspectral classification
14	[68]	SVM, RF, LR, KNN and DT	Indian pines	99.65	PCA and MNF techniques were used to minimize the noise and unnecessary bands present in the dataset. The efficacy of the models was evaluated by considering different performance measures such as confusion matrix, overall accuracy, and training time
15	[69]	Combination of CNN and random forest (RF)	KSC, Indian Pines	98.32	The approach used has successfully extracted spatial-spectral information from the HSI data.
16	[70]	Deep Random Forest	Indian Pines, University of Pavia, Salinas	99.91	The proposed method first uses MAPC as the feature for a sample. Secondly, the deep random forest offers the benefits of hierarchical feature extraction similar to a neural network but without the computationally expensive training process that requires much training data and time
17	[71]	K-NN	Indian Pines	81.6	A new similarity measure to compare the retained Fourier Descriptors of the candidate pixel with those of the reference pixel.
18	[72]	K-NN	Salinas, Indian Pines	91	With the development of hyperspectral imaging technology, precise and efficient object detection and classification methods are increasingly essential.

Table 3 (continued)

S. No	Ref	Method	Dataset	Metrics (OA)	Remarks
19	[73]	maximum likelihood (ML)	Indian Pines, HYDICE	85.52	The proposed CAP-MLC method improves the accuracy of Maximum Likelihood classified hyperspectral images, even with limited training samples. It also clusters based on collected classes.
18	[74]	SAM	Real data	90	Major spectral differences were observed around 975, 1215, and 1450 nm
20	[75]	SAM, SVM, RF	Dioni, Loukia	92	The SVM and RF algorithms significantly underperformed the SAM regarding classification accuracy. RF had a slight edge over SVM.
21	[76]	SAM	Pavia University, Salinas, Botswana	84.47	To enhance the classification accuracy, further research is needed.
22	[77]	Spectral Angle Mapper (FSAM)-Active Learning (AL)	Salinas, KSC, Pavia University	99.52	It has been observed that the use of reference spectra in SAM for HSI classification, which is assumed to represent pure spectra, may sometimes result in spectral mixture problems, particularly in cases of low or medium spatial resolution.
23	[78]	LSTM	Pavia University, Salinas	96.20	To further advance this research, it may be worthwhile to consider incorporating these methodologies within a temporal context, with continued emphasis on the encoding of spatial contextual information.



**Table 4** Literature survey of existing methods based on unsupervised classification techniques for hyperspectral image classification

S. No	Ref	Method	Dataset	Metrics (OA)	Remarks
1	[79]	Massive Self-organizing Maps	Indian pines, Salinas	95.95	MSOM map parallelization work is still pending. And to justify the Massive SOM theoretical investigative foundations are still pending, to get a better classification performance.
2	[80]	PCA	Indian Pines, University of Pavia	95.95	Automatically determining the optimal number of superpixels for segmentation is vital in modern image analysis. The traditional approach of trial and error is not only time-consuming but can also yield inaccurate results.
3	[81]	CNN	Indian Pines, KSC, Salinas	97.47	To reduce the manual intervention required for network design, further study will be implemented on optimizing hyperparameters in the proposed networks.
4	[82]	3D convolutional autoencoders	Salinas, Pavia University, Indian Pines, and real-time data	89.5	The proposed method can be computationally expensive.
5	[83]	CNN	Indian pines, Salinas, KSC, Botswana	98.09	Extend the network by incorporating other layers, such as depth-wise convolution, to enhance accuracy and performance.
6	[84]	K-means and density peaks (DP) clustering	Real data	92.70	This approach has the potential to differentiate between pathological sites in human small intestinal tissue, which could be of significant assistance to clinicians in the accurate diagnosis of small intestine tissue.
7	[85]	local neighborhood structure preserving embedding (LNSPE)	Pavia University, Botswana	94.23	Combining the priori label information to improve the classification performance of LNSPE further.
8	[86]	Subpixel component analysis (SCA)	Indian Pines, KSC, Pavia University	97.79	The proposed SCA technique needs to be implemented in high-performance computing systems.
9	[87]	Random Subspace (RS)-based KNCCRT ensemble framework	Salinas, Pavia University, Indian Pines	94.37	When training with extensive data, the number of bands in a feature subset must be pre-defined.

**Table 4** (continued)

S. No	Ref	Method	Dataset	Metrics (OA)	Remarks
10	[88]	PLG-KELM and PCA, and LBP	Indian pines, Houston, Pavia university	99.10	The proposed algorithm has a lower operation efficiency.
11	[89]	A novel multiscale residual network (MSRN)	Salinas, Pavia University, Pavia centre	98.69	Experimental results are compared with the three-benchmark HSIC, but the proposed method archives good results.
12	[90]	A novel framework to construct an ELM ensemble model	Salinas Pavia University Indian Pines Houston	94.84	The anticipated technique delivers the inexpensive results The MVRR-EP ensemble pruning approach does not guarantee a globally optimal result.

**Table 5** Existing methods based on hyperspectral image classification using semi-supervised classification algorithms were reviewed

S. No	Ref	Method	Dataset	Metrics (OA)	Remarks
1	[91]	Graph-based methods	Indian Pines	49.69	The presence of numerous unlabeled samples in the image, coupled with the integration of contextual information, can significantly aid in their interpretation. By leveraging the contextual information, it is possible to more accurately decipher the abundance of these unlabeled samples
2	[92]	SVM, gradient boosting, Gaussian classifier, and linear perceptron	Jasper HSI	100	The AWS cloud server demonstrated a run time of approximately 6 h and 47 min, while the same process on a regular PC would require a week for training and classification
3	[93]	graph convolution networks (GCNs)	Indian Pines, Salinas, Pavia University	96.87	The proposed framework has demonstrated exceptional performance surpassing that of the current state-of-the-art methods on three different hyperspectral image datasets. This outcome suggests that the framework can be considered a viable solution to enhance the quality of hyperspectral image analysis
4	[94]	NearPseudo, pseudo-labels	Xiongan	76.2	Analyzing the impact of NearPseudo on the classification accuracy of standard classifiers
5	[95]	semi-supervised CD with multilayer cascade screening strategy (MCS4CD)	Hermiston, USA Dataset, River, China	97.84	The CD method uses nonlinear transformation and spatial constraints to classify HSI datasets with limited labeled samples
6	[96]	Self-training	Indian Pines, Pavia University, Salinas	97.33	The issue of boundaries between different scenes arises when adjacent samples have different classes
7	[97]	two branch autoencoder (TBAE)	Indian Pines, Pavia University,	92.59	The proposed approach outperforms CNN and ANN with small sample sizes. Further combinations of training can improve accuracy

Table 5 (continued)

S. No	Ref	Method	Dataset	Metrics (OA)	Remarks
8	[98]	Low-rank Laplacian regularizer	Salinas, Indian Pines, Pavia University	85.90	The proposed technique achieves extraordinarily well when only an insignificant number of labelled examples are accessible
9	[99]	A lightweight tensor attention-driven ConvLSTM neural network (TACLNN)	Pavia University, Salinas, Indian Pines	98.25	The proposed technique reduces parameters and storage while improving accuracy
10	[100]	Mathematical morphology (Mm) post-processing	Indian pines	93.74	The proposed method achieves good results for post-processing hyperspectral image classification
11	[101]	Adaptive DropBlock-Enhanced Generative Adversarial Networks	Salinas, Pavia University, Indian pines	95.38	To improve classification accuracy, different types of self-attention networks are available
12	[102]	GPU-Cat Boost, the GPU-accelerated Cat Boost-Forest	Pavia University, Grass-DFC2013, Grass-DFC2018	95.75	Compared to existing decision-tree-based algorithms, the suggested method provides better outcomes while retaining high computing efficiency
13	[103]	A novel structural-kernel collaborative representation (SKCR)	Indian Pines, Washington DC	99.36	The new technique achieves better results compared to other models.

**Table 6** Existing methods based on hyperspectral image classification using attention-based classification models were given

S. No	Ref	Method	Dataset	Metrics (OA)	Remarks
1	[104]	Attention-Based Adaptive Spectral–Spatial Kernel ResNet	Indian Pines, KSC, Pavia University	99.86	After rigorous testing, the A2S2K-ResNet algorithm has been found to demonstrate superior performance compared to other state-of-the-art methods on HSI datasets, even when training samples are limited
2	[105]	3-D channel and spatial attention-based multiscale spatial–spectral residual network (CSMS-SSRN)	Indian Pines, Pavia University, KSC	99.96	The CSMS-SSRN model makes use of deep multiscale features along with a general attention mechanism that can be conveniently adapted to address various remote sensing classification tasks
3	[106]	Spectral-Spatial-Wise Attention-based Siamese Network	Bay Area, Santa Barbara, River, Farmland	98.77	The experimental results obtained from four hyperspectral images unequivocally demonstrate that the proposed method outperforms all state-of-the-art approaches in terms of accuracy for change detection
5	[107]	Attention-Based Pyramid Network	Vaihingen, Potsdam, Indian Pines, and Pavia University	90.91	To encourage the effective implementation of deep learning techniques in the perception of remote sensing images
6	[108]	Attention-Based Domain Adaptation Using Residual Network	Pavia City	97.24	The transferred parameters should be investigated to enhance classification performance in DA
7	[109]	multiscale attention-based hybrid spectral network and UNet (MSA-HybridSN-U)	Indian Pines, University of Pavia, Houston	99.97	This study makes a significant contribution to the development of hyperspectral image classification models

**Table 7** Existing methods based on hyperspectral image classification using CNN-based classification models were given

S. No	Ref	Method	Dataset	Metrics (OA)	Remarks
1	[110]	CNN	Salinas Valley, Indian Pines,	98.1	Possible misclassification of similar labels can be resolved by augmenting input vectors with more spatial and spectral features from hyperspectral imaging data
2	[111]	2D-3D CNN	Indian Pines, Salinas, KSC	98.33	The process of refining features in the Convolutional Neural Network (CNN) for Hyperspectral Image Classification (HSIC) heavily depends on the availability of band information
3	[112]	A Hybrid 3D-2D CNN	Salinas, Pavia University, Indian Pines	99.17	The superiority of the proposed method was confirmed through experiments conducted on three datasets and compared against three classification methods. The study found that even with limited training samples and noise, the proposed method outperformed the other classification methods
4	[113]	Multidimensional CNN combined with an Attention Mechanism	Salinas, WHU, Pavia University	97.34	The CBMS attention module has been shown to greatly enhance the accuracy of HSI classification. I have checked the text for any errors and corrected them
5	[114]	Attention Aided CNNs	Houston 2013, Houston 2018, HyRANK	90.38	The integrated model of both sub-networks can further enhance the performance
6	[115]	3D 2D CNN	Indian Pines, KSC, University of Pavia, Salinas Scene	99.92	The training and testing time on IP and KSC has been significantly reduced when using only the fast convolution block
7	[116]	Pixel cluster CNN and spectral-spatial fusion (SSF)	Pavia University, Indian Pines, Salinas	94.02	The CNN architecture extracts essential features that can be employed as classification rules

Table 7 (continued)

S. No	Ref	Method	Dataset	Metrics (OA)	Remarks
8	[117]	2D-3D CNN	University of Pavia, Indian Pines, Salinas, Botswana	99.94	The proposed method utilizes a more concise neural network model to improve generalization while maintaining classification accuracy across datasets
9	[118]	Hybrid CNN	Bloodstain	96	The proposed method will help train deep models without the need for extensive data augmentation, thus improving the efficiency and effectiveness of the overall model
10	[119]	Mixed Convolutions and Covariance Pooling	Indian Pines, University of Pavia, and Salinas Scene	100	This will involve exploring various optimization techniques to fine-tune the model's weights, thus improving its performance and versatility
11	[120]	CNN	Real-time data	99.3	The results indicate that Deep learning CNN can effectively work with small datasets, making it a suitable choice for applications with limited data
12	[121]	Convolutional Neural Networks (CNNs)	Real-time data	98	The findings provide valuable insights for researchers in this field and lay the groundwork for future research on pest infection, offering new ideas and avenues for further exploration



**Table 8** Existing methods based on hyperspectral image Prediction

S.No	Ref	Objective	Method	Remarks
1	[122]	To estimate and predict the corn seeds	SVM, ELM, and IDCNN	Improving seed recognition models for germination based on temporal sequences is critical
2	[123]	To assess five dissimilar scanning approaches to their possible use in potato valuation and sorting	Dielectric and LF-NMR	The proposed method is unsuitable for establishing analytical methods for robust prediction models
3	[124]	To pretend the spatio-temporal dynamics of land use and land cover (LULC) alterations in Thimphu city	CA-Markov	The proposed method predicts significant changes in Thimphu city
4	[125]	To predict the land use and land cover (LULC) changes in Tuy Duc districts	CA-Markov	The suggested approach predicted the LULC change in the forest
5	[126]	To forecast the trajectory of LULC changes in the province of Dak Nong over time	RF, MC with Multilayer Perceptron Neural Network (MLP-NN)	The suggested method was employed to project changes in land use and land cover. The findings from this study can aid in developing sustainable land management strategies and policies
6	[127]	To meet the goals of maintainable expansion over the area	MCM using MLPNN	The proposed model still needs to be revised to predict the socioeconomic data and government policy
7	[128]	To know the agriculture growth in the location of Cerrado/ Atlantic Forest ecotone's native vegetation areas	CA and a Markov forecast model	The suggested approach successfully pretended the development of agricultural detections located in the Prata River basin
8	[129]	To determine whether it is feasible to estimate and forecast certain vital signs. Seeds from Quercus variabilis	Savitky-Golay first derivative (SG 1st) and multiple scatter correction (MSC)	The technique proposed here confidently enables non-destructive valuation of <i>Quercus variabilis</i> seed vitality and precise prediction in a short duration
9	[130]	To identify the delicate spectral and textural traits, including information about mangrove pests and diseases	Successive projection algorithm (SPA), random forest (RF)	The proposed technique is an excellent possibility for defining leaf features' spatial (three-dimensional) distribution under different pest and disease violence

an innovative hybrid classifier for hyperspectral images that employs the Bat Algorithm (BA) to optimize the Convolutional Neural Network (CNN) architecture. The resulting BAT-CNN classifier is tested on three different hyperspectral datasets and has demonstrated superior accuracy compared to the standalone CNN classifier. This new approach shows great promise in advancing the classification of hyperspectral images and has the potential to significantly improve the accuracy of remote sensing applications.

In [45], authors proposed CM-CNN, a new 3D convolutional neural network, for hyperspectral image classification. CM-CNN achieved a stable Kappa coefficient, confusion matrix accuracy above 95%, and almost no obvious classification errors. A two-stage learning algorithm for the classification of hyperspectral images was proposed in [46]. The algorithm optimizes classification results through Kernel Singular Value Decomposition-Multiple Kernel learning and Conditional Random Field. The authors in [47], proposed a novel hyperspectral image classification algorithm and introduced a hyperspectral sky imaging dataset. The dataset is augmented using multiple clustering, leading to higher pixel classification accuracy. Gradient boosting methods outperformed benchmark algorithms. A complementary Integrated Transformer Network (CITNet) for the classification of hyperspectral images in [48]. After conducting experiments with CITNet, it was found that the utilization of Conv3D and Conv2D to extract shallow semantic information, along with the implementation of a channel Gaussian modulation attention module to enhance secondary features, led to improved classification performance. The present study introduces a novel algorithm, namely Class Information-based Principal Component Analysis (CI-PCA), that aims to enhance hyperspectral image classification [49]. To create a CI-PCA image, we need to choose particular pixels or regions for training purposes for each defined class. Afterward, we compute the Principal Component Analysis (PCA) for each class's training data individually. Finally, we merge the PCA results of each class to form the ultimate CI-PCA image. The efficacy of this method has been proven with the utilization of two genuine hyperspectral datasets. In [50], a novel approach to vegetation classification has been introduced, called DCKELM-SPATIAL. This technique utilizes a deep composite kernel extreme learning machine that leverages spatial feature extraction, employing the Gabor filter and super-pixel density peak clustering method to generate a fresh set of spatial composite kernels. Empirical findings have demonstrated that this strategy surpasses numerous traditional and sophisticated methods in terms of classification precision.

The authors in [51], To address the common concerns of overfitting and excessive parameters in deep learning models for hyperspectral image classification, a team of researchers has proposed the Hybrid Fully Connected Tensorized Compression Network (HybridF-TCN). This innovative network has been shown to achieve state-of-the-art classification performance with a minimal number of parameters. An advanced K-means hyperspectral classification algorithm that confidently portrays the significance of bands through variance coefficients and integrates inter-class information to achieve optimal clustering at a global level is proposed in [52]. A spectral-spatial classification method for homogeneous regions in hyperspectral images, based on locality-constrained joint-sparse and weighted low-rank, to enhance classification accuracy [53]. Other classification methods are outperformed by this one in terms of accuracy. In [54], through the application of self-supervised masked image reconstruction, the researchers have successfully improved transformer models for hyperspectral remote sensing imagery. Their insightful findings indicate that modifying the architecture of the vision transformer can yield significant enhancements in the accuracy of land cover classification tasks. It is noteworthy that the transformer model surpasses randomly initialized transformers and 3D convolutional neural networks by an impressive 7–8%, even when only a mere 0.1–10% of the training labels are accessible.

Hyperspectral image classification has proven to be an invaluable tool in categorizing land use and cover, contributing greatly to the fields of land management, urban planning, and environmental studies. The literature has discovered that machine learning and deep learning approaches such as SVM, CNN, RF, SOM, KNN, SAM, LSTM, and Attention models are the most effective methods for carrying out this classification process. These methods have been found to have the highest accuracy, consistency, and robustness. They can handle complex data sets and distinguish between different spectral signatures of land cover types. Figure 6 shows the taxonomy of hyperspectral image classification and Fig. 7 shows the existing models comparison for hyperspectral image classification.

### 3 Applications of hyperspectral imaging

According to [131], Hyperspectral imaging technology is utilized in various applications to solve real-world problems. Hyperspectral imaging is a remote sensing technique that captures data across numerous narrow, contiguous wavelength bands. Thanks to its high spectral resolution, researchers can identify and classify objects with exceptional precision. This technology has a wide range of applications, including:

1. **Agriculture:** HSI classification finds application in various agricultural areas such as crop mapping, pest and disease detection, stress detection, and yield estimation. Crop mapping using HSI helps to create accurate maps of different crop types like corn, soybeans, wheat, etc. This information can be utilized to manage crops more effectively and increase yields. HSI can detect and identify pests and diseases in crops, which helps farmers take early measures to control them before they cause significant damage. HSI can also help to detect stressed crops, for instance, crops facing drought, nutrient deficiencies, or salinity. This knowledge can assist farmers in identifying and addressing issues before they lead to reduced yields. Finally, HSI can estimate crop yields, which can be used to plan for harvest and market crops more effectively [1].
2. **Seed viability study:** Hyperspectral imaging (HSI) is a valuable tool to study seed variety in several ways. One potential benefit of utilizing spectral signatures is the ability to identify different seed varieties accurately. This can aid in ensuring seed quality and selecting the most suitable seed variety for planting purposes. Another one is, that HSI can be used to assess the quality of seeds by analyzing their spectral signatures. This enables the identification of damaged, diseased, or immature seeds. Lastly, HSI can predict seed germination by analyzing their spectral signatures, which is helpful for crop yield optimization and planning planting [132]. As HSI sensors are becoming increasingly affordable and portable, they are expected to be used in more innovative and exciting ways. HSI is a powerful tool for studying seed variety and can potentially revolutionize the farming industry [133] (Fig. 8). For Example, the Corn seed monitoring using hyperspectral Imaging is shown in Fig. 8 (a) Original Image, Fig. 8 (b) Partial least squares discriminant analysis, Fig. 8 (c) Binary Image.
3. **Biotechnology:** HSI is a versatile technology employed in several ways in the biotechnology field. Some key areas where HSI is being utilized are drug discovery, disease diagnosis, and monitoring. HSI can be used to diagnose and monitor various diseases, including cancer, Alzheimer's disease, and Parkinson's disease. It can also identify tumors and track their growth over time. [135].

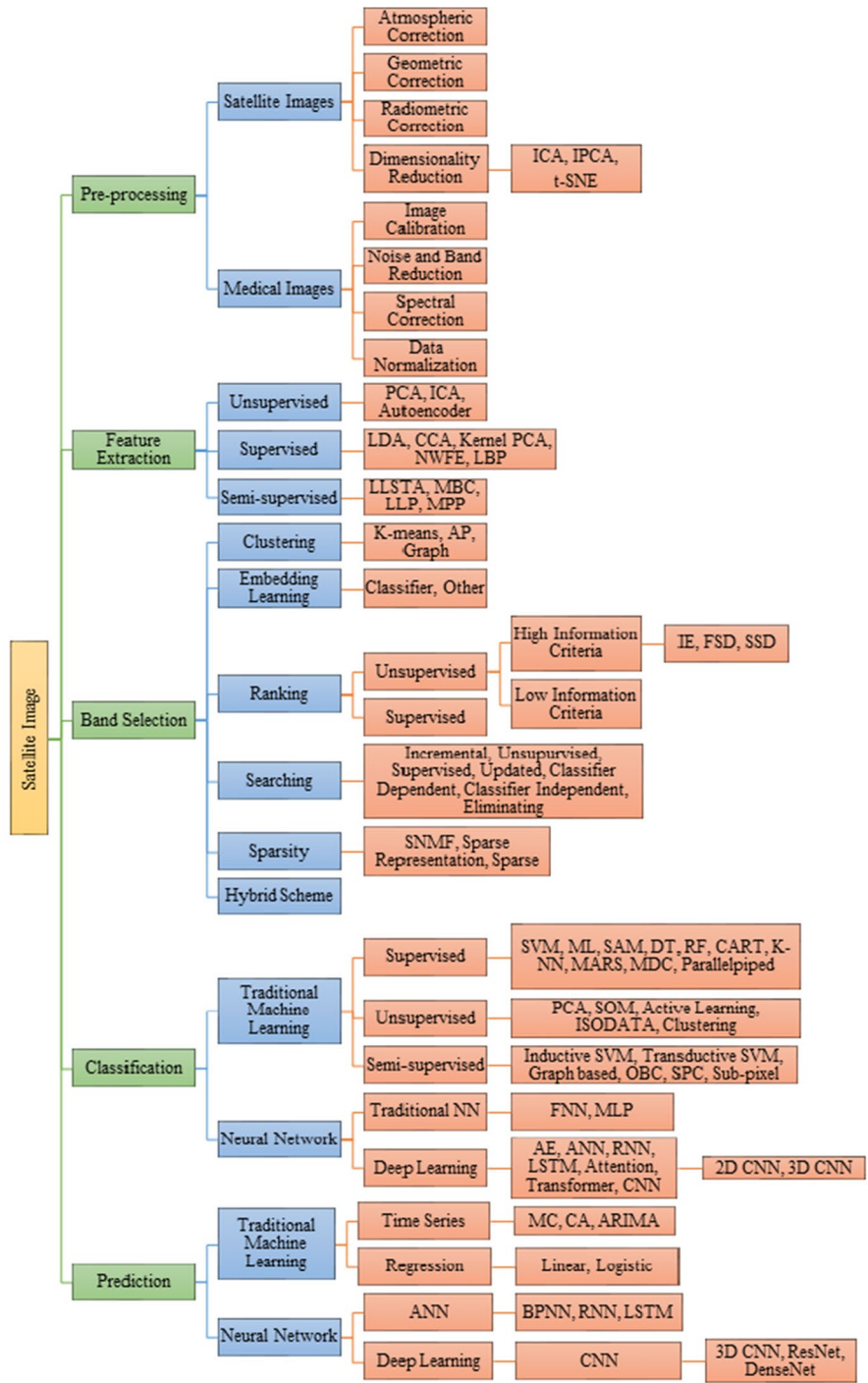
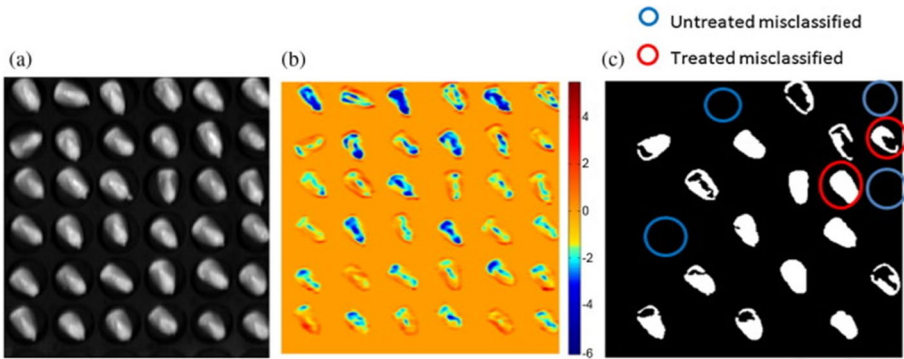
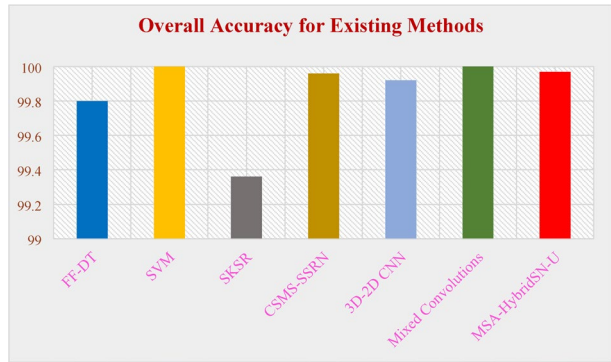


Fig. 6 The taxonomy of hyperspectral image classification

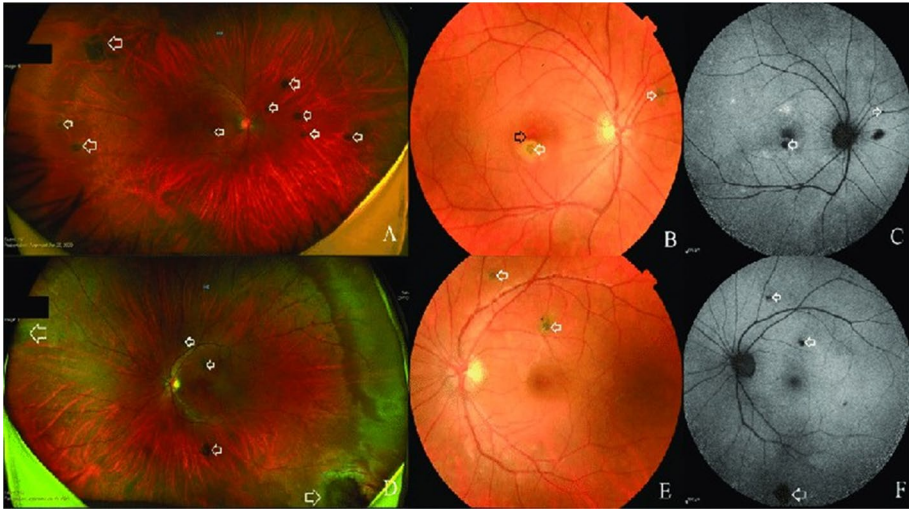
**Fig. 7** Existing model comparison for hyperspectral image classification



**Fig. 8** Corn seed monitoring using hyperspectral Imaging **(a)** Original Image, **(b)** Partial least squares discriminant analysis, **(c)** Binary Image [134]

4. **Eye Care:** Eye care is an emerging field that utilizes hyperspectral imaging technology to improve eye disease detection, diagnosis, and treatment. According to [136], doctors can accurately identify and classify objects due to their high spectral resolution. Early detection of subtle changes in the eye is crucial to prevent vision loss caused by conditions like glaucoma, diabetic retinopathy, and age-related macular degeneration. HSI is used to track the progression of eye diseases over time, enabling doctors to adjust treatment plans and monitor the effectiveness of treatment. It can also guide surgeons during ophthalmic procedures, such as cataract surgery and retinal detachment repair, which can improve the accuracy and safety of surgery. Figure 9 shows the Fundus examination of both eyes and it was documented with ultrawide field imaging, color fundus photography, and fundus autofluorescence imaging using hyperspectral imaging.
5. **Food Processing:** Hyperspectral imaging technology is used for food processing to enhance food processing operations' safety, quality, and efficiency. According to [137], Hyperspectral imaging can detect foreign objects, such as metal, plastic, and glass, in food products, improving food safety and preventing consumers from getting sick. It can assess the quality of food products by measuring factors such as ripeness, freshness, nutritional content, and chemical composition. This information can improve the quality of food products, reduce waste, and ensure food safety. Also, Hyperspectral

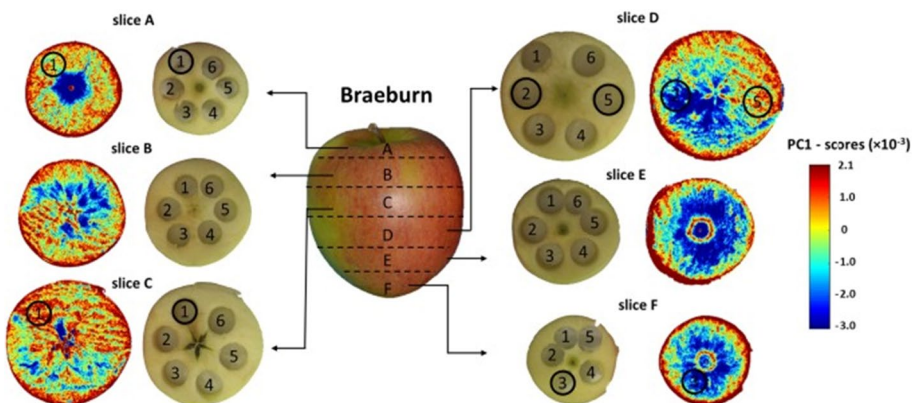




**Fig. 9** Fundus examination of both eyes, documented with ultrawide field imaging, color fundus photography, and fundus autofluorescence imaging [136]

imaging can sort and grade food products based on their quality and other characteristics, improving the efficiency of food processing and ensuring that consumers receive high-quality products (Fig. 10).

6. **Environmental Monitoring:** Using hyperspectral imaging, using hyperspectral imaging technology to monitor the environment for changes or disturbances. According to [138], Hyperspectral imaging can be used to monitor various environmental factors such as air and water quality, land cover, vegetation and soil health, geological hazards, natural disasters, and climate change.
7. **Forensic Science:** In [139], Forensic science using hyperspectral imaging involves using hyperspectral imaging technology to collect and analyze evidence from crime scenes and other forensic settings. To detect the bloodstains, even in low-light conditions or on dark fabrics. HSI can enhance the visibility of fingerprints, even on



**Fig. 10** Sliced Braeburn apple detection [137]

complex surfaces like metal or plastic. It can also detect gunshot residue and other evidence related to firearms. Hyperspectral imaging can detect and identify drugs and explosives, even in trace amounts. To analyze fibers from clothing or other materials, which can help to link suspects to crime scenes. HSI can detect documents, such as handwriting and ink analysis, to provide substantial evidence in forgery or fraud cases [140] (Fig. 11).

8. **Thin Films:** Thin films in hyperspectral imaging refer to thin layers of materials that possess unique optical properties that can be used to manipulate light in various ways, such as reflecting, transmitting, or absorbing light at specific wavelengths [142]. HSI is used to create filters that can be used to select or block specific wavelengths of light. This can be useful for applications such as hyperspectral microscopy, where it is crucial to image specific chemical species or structures. Thin films can coat optical components, such as lenses and mirrors, to enhance performance. For instance, anti-reflective coatings can be used to minimize glare and improve image quality. Also, thin films can be used to create sensors that can detect and measure specific wavelengths of light. This can be useful for environmental monitoring and food safety inspection [143].
9. **Oil and Gases:** Hyperspectral Imaging (HSI) effectively detects, identifies, and quantifies oil and gas in various environments. HSI can directly detect oil and gas by identifying their unique spectral signatures. For example, oil and gas have characteristic absorption bands in the infrared region of the spectrum. HSI can indirectly detect oil and gas by identifying features associated with oil and gas deposits. HSI works by analyzing reflected light from the surface of the earth at many different wavelengths, creating a detailed spectral signature that can help to identify subtle changes in vegetation, surface temperature, and other environmental factors that may be indicative of the presence of oil and gas deposits. HSI can also be used in various oil and gas exploration and production applications such as mapping oil spills, monitoring pipeline leaks, and identifying potential drilling locations [144].

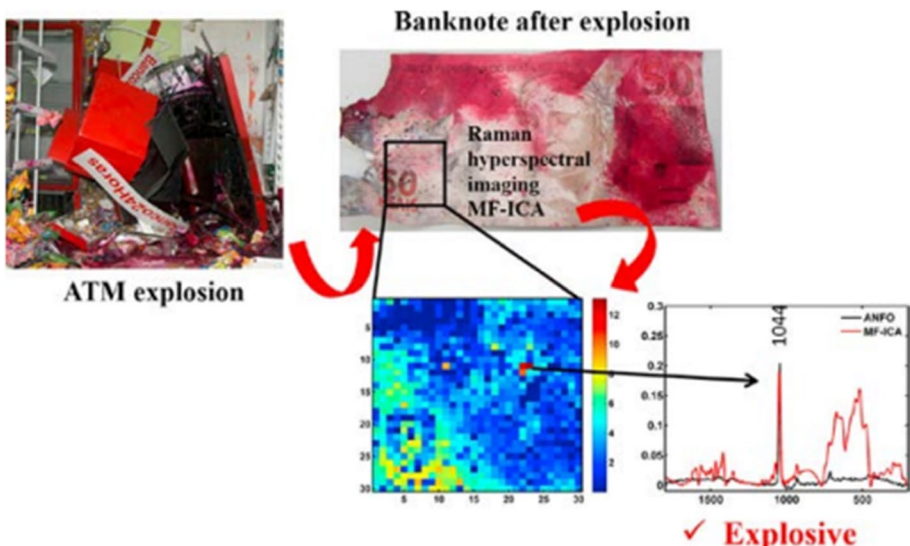


Fig. 11 Representation of ATM explosion and its spectral signatures [141]



10. **Cancer Diagnosis:** Hyperspectral Imaging (HSI) detects, identifies, and characterizes cancer cells and tissues. HSI can detect and identify cancer cells and tissues in various ways, to detect cancer cells and tissues by identifying their unique spectral signatures. For instance, cancer cells often have different absorption and scattering properties than normal cells. HSI can indirectly detect cancer cells and tissues by identifying features that are associated with cancer. For example, HSI can be used to identify changes in blood flow, tissue oxygenation, and other cellular and physiological processes that can be indicative of cancer. HSI has various applications in cancer diagnosis, such as identifying cancerous tissue margins during surgery, monitoring tumor progression and treatment response, and identifying biomarkers for early cancer detection [145]. Hyperspectral imaging for medical applications are represented in (Fig. 12).
11. **Animal Detection:** Apart from the applications discussed above, animal detection applications are also most attractive, which use various input modalities, namely RGB image [145, 147], thermal image [148], unmanned aerial images [149], unmanned ground images [150, 151], and hyperspectral images. Hyperspectral imaging uses are increasing daily, not only in animal detection and classification but also for animal food quality, remote animal health monitoring, and animal disease detection etc., HSI is also used in the poultry form sector.

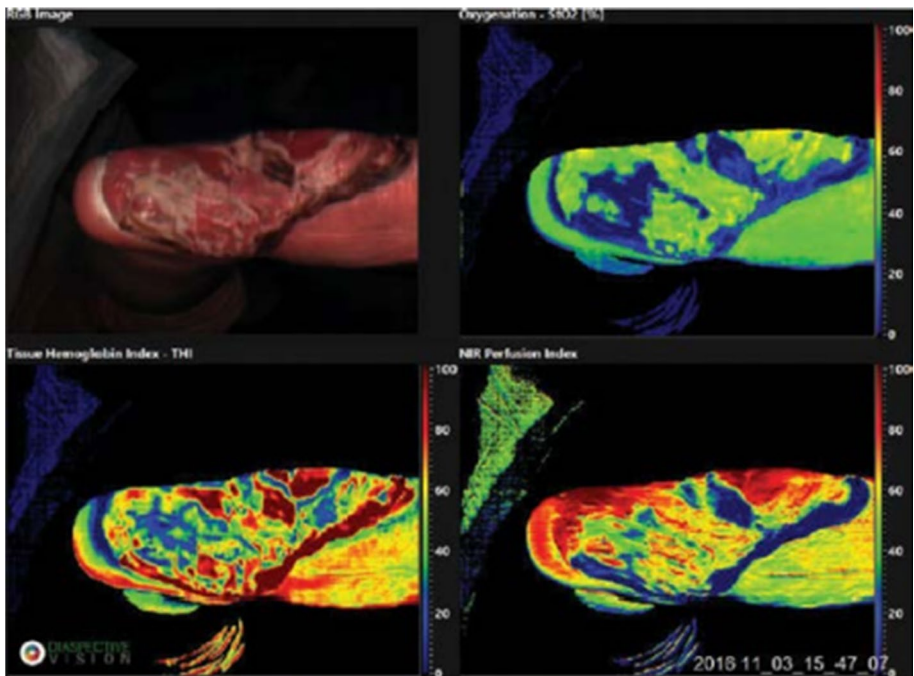


Fig. 12 Hyperspectral imaging in medical applications [146]

## 4 Pre-processing

Hyperspectral image pre-processing is a crucial step in any hyperspectral imaging workflow, as it prepares the images for further analysis and enhances the accuracy and reliability of the results. Hyperspectral imaging also requires pre-processing, which includes removing noise, correcting atmospheric effects, and reducing data dimensionality. In hyperspectral image pre-processing, we have separate techniques for medical and satellite images. Fig. 13 details the pre-processing techniques involved in hyperspectral imaging.

1. **Pre-processing Techniques for Hyperspectral Satellite Images:** Hyperspectral satellite image pre-processing, Atmospheric Correction (AC), Radiometric Correction (RC), Geometric Correction (GC), and Dimensionality Reduction (DR) are used.
  - a. **Atmospheric Correction (AC):** Atmospheric correction is a crucial process that eliminates the atmospheric effects on the spectral signatures of materials in an

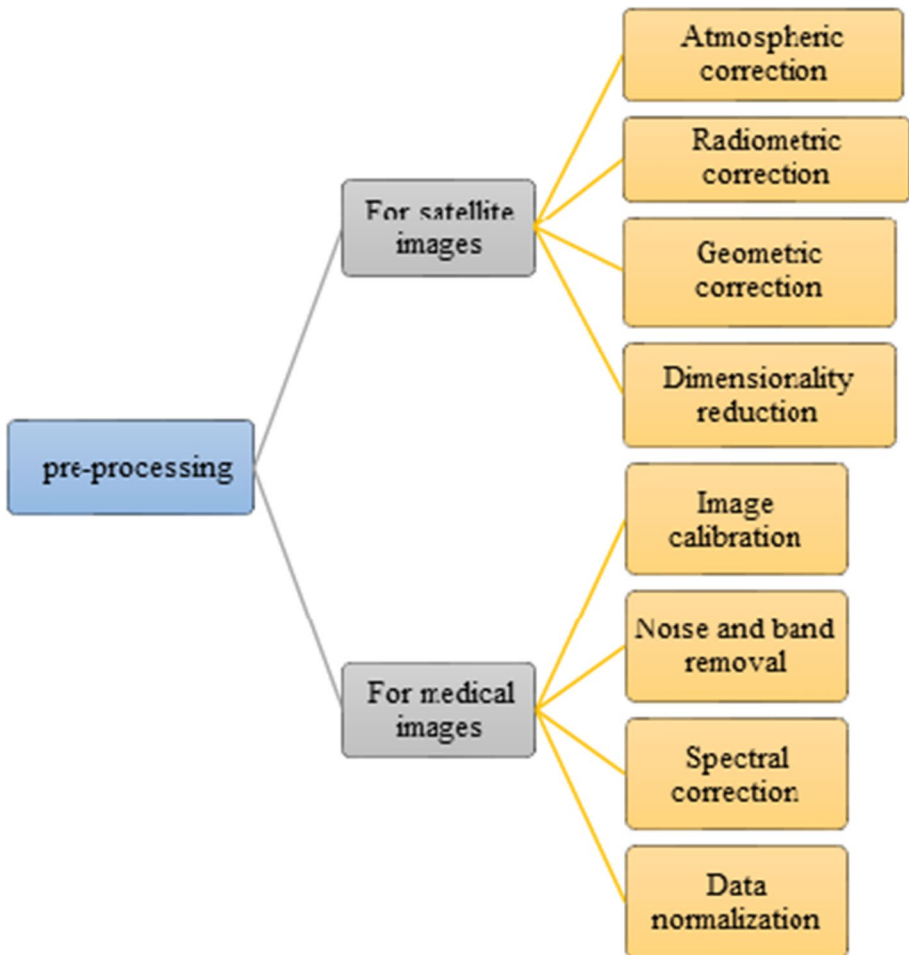


Fig. 13 Pre-processing techniques

image. Atmospheric gases and particles can distort spectral signatures in images through light absorption and scattering. To ensure the accuracy and reliability of remote sensing data, it is crucial to understand atmospheric conditions and apply appropriate correction techniques. Accurate identification and classification of materials in the image is essential for hyperspectral imaging, and atmospheric correction ensures that researchers can achieve this without any hindrance [152]. In hyperspectral imaging, several techniques can be used for atmospheric correction. One commonly used method is the Radiative Transfer Model (RTM). An RTM is a computer model replicating the light path through the atmosphere. The model estimates the amount of light absorbed and scattered by the atmosphere, allowing researchers to eliminate these effects from the image. Another method used for atmospheric correction in hyperspectral imaging is to use a reference spectrum. A reference spectrum is a spectral signature of a known material, such as a white reference panel. Researchers compare the spectral signature of a pixel in the image to the reference spectrum to estimate the amount of light absorbed and scattered by the atmosphere and remove these effects from the image [153].

Atmospheric Correction (AC) is used for quantitatively estimating the water surface reflectance. Also, it can be utilized to determine the absorptions of the numerous water constituents concluded by an inversion algorithm. The AC's aerosol types are enlarged from 12 to 80 [5]. The wavelengths used for the AC are when the water leaving radiance is expected to be zero. It has remained the designated near-infrared meant at comparatively clear waters, and the shortwave infrared designed for exceedingly turbid and SWIR bands is not available. AC is a vast source of error correction in hyperspectral remote sensing, with the external assets even lower than perfect circumstances [154]. Also, Fig 14 displays the improved repossessions with contiguousness correction shown in red, the improved repossessions starved of adjacency correction in orange, and the standard retrievals in blue. The light grey spectra represent the flying levels.

- b. **Radiometric Correction (GC):** Radiometric correction, a crucial step in hyperspectral imaging, transforms Raw Digital Numbers (DNs) of image pixels into physical units, such as radiance or reflectance, inspiring a deeper understanding of the imaging process. The RDNs of the image pixels are influenced by several factors, including the sensor gain and offset, the solar irradiance, and the atmospheric conditions [155]. Radiometric correction in hyperspectral imaging can be done through various methods. One widely used method is to employ a calibration equation. This equation establishes a relationship between the RDN values of the pixels in an image and the radiance or reflectance of the materials in the image. A reference panel with a known radiance or reflectance is typically used to derive the calibration equation [156]. It has been observed that SAR data calibration is an effective method of obtaining essential radar backscatter information for RC. To resolve the Digital Elevation Model (DEM) for the radar backscatter image, the superiority of any radiometric terrain correction is massively dependent [157]. Also, Fig. 15 (a) shows the original image, and (b) gives the radiometrically corrected image of the AHI (Airborne Hyperspectral Imager) made in China. It is a non-combatant remote sensor; this will focus more on the quantifiable reclamation of surface geophysical constraints. The AHI sensor has a spectral range of 400 to 1000 nm.

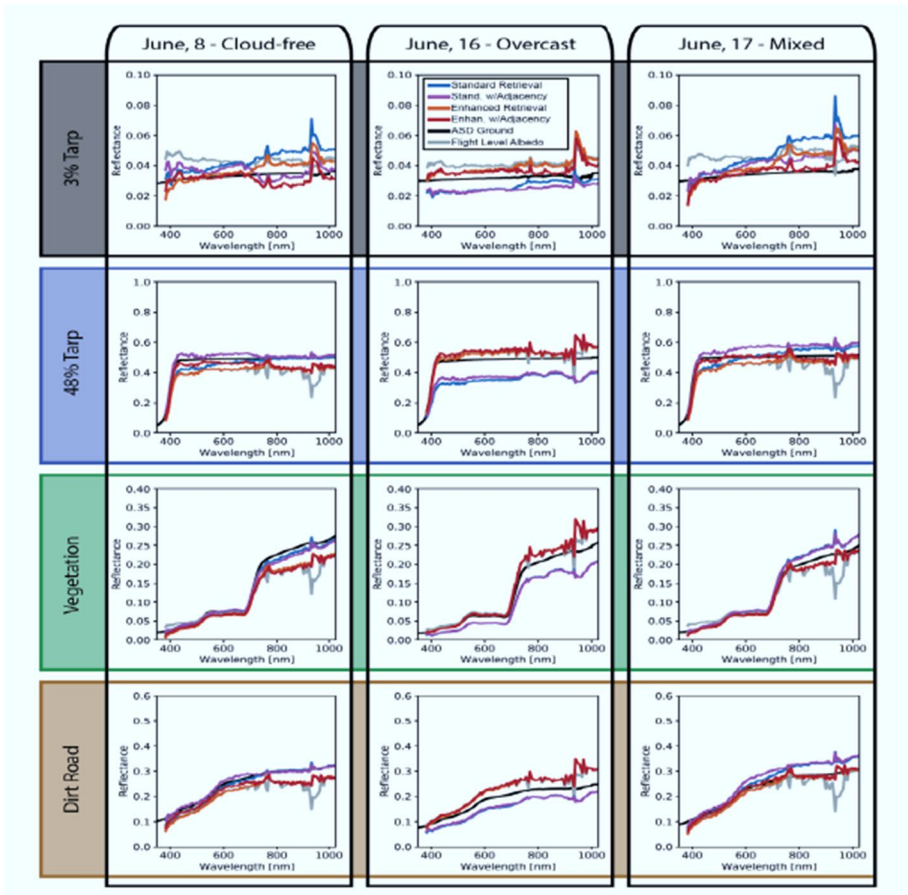


Fig. 14 Results of Atmospheric correction [154]

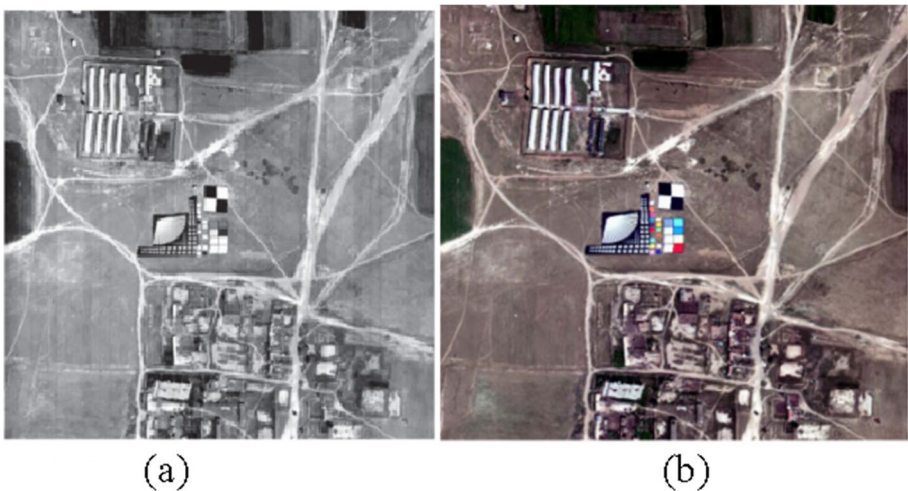


Fig. 15 a Original image, b Corrected image of radiometric correction [157]

- c. **Geometric Correction (GC):** Geometric correction is a process used in hyperspectral imaging to rectify geometric distortions in the image. These distortions may arise due to various factors, such as the sensor platform, the Earth's curvature and rotation, and atmospheric refraction. Hyperspectral imaging relies on geometric correction to ensure the hyperspectral image aligns accurately with other geospatial data, such as satellite imagery and maps [158]. It decreases the terrain displacement and helps advance the image's positional accuracy. It is a technique to correct an image's errors and increase quality [159]. Also, Fig. 16. Represents the difference between before and after geometric correction.
- d. **Dimensionality Reduction (DR):** While hyperspectral imaging requires a careful balance between spectral band reduction and the retention of crucial information, dimensionality reduction techniques can be utilized to achieve this objective effectively. Hyperspectral images often contain hundreds or thousands of spectral bands, making them computationally challenging to analyze [160]. Therefore, dimensionality reduction is essential to simplify the data and facilitate efficient analysis. In dimensionality reduction, we have the following methods: IPCA, ICA, and t-SNE etc., Independent Component Analysis (ICA) is a powerful statistical technique that efficiently segregates data into its independent components—signals that are impervious to further simplification. ICA is commonly utilized in hyperspectral imaging to extract essential features hidden within the data [161]. Incremental Principal Component Analysis (IPCA) is a technique for reducing the dimensionality of a dataset. The technique allows for the updating of principal components as new data is added. This is different from traditional batch Principal Component Analysis (PCA), which calculates all principal components of a dataset together. IPCA is useful when dealing with large datasets that cannot be processed simultaneously. It enables the calculation of principal components using smaller subsets of data, thereby reducing memory requirements and computational complexity [162]. t-SNE (t-distributed Stochastic Neighbour Embedding) is another dimensionality reduc-

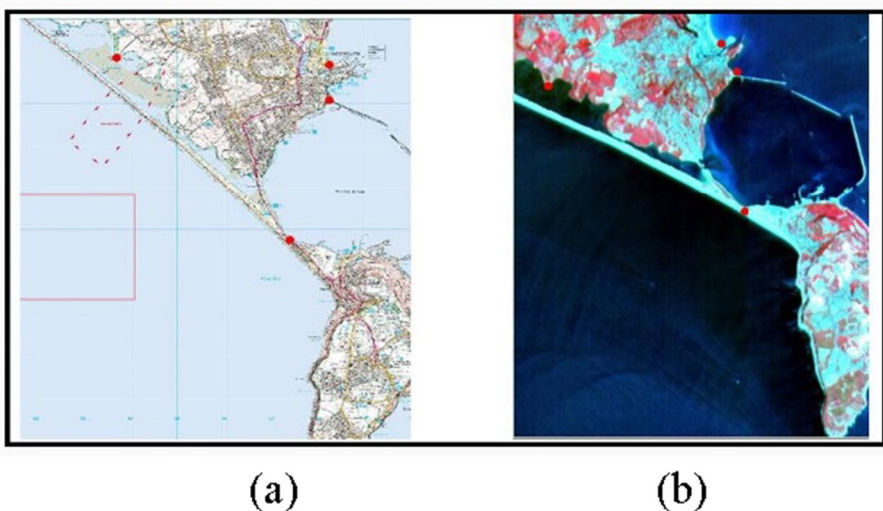


Fig. 16 Geometric correction (a) Before and (b) After geometric corrections [159]

tion technique commonly used for data visualization. Visualizing high-dimensional datasets can be a complex task. Fortunately, there is a technique that can simplify this process. t-SNE is a valuable tool that reduces high-dimensional datasets into a low-dimensional space, such as a 2D or 3D space, making it easier to visualize. By mapping similar data points in high-dimensional space to nearby points in low-dimensional space, while mapping dissimilar data points to distant points, t-SNE allows for a more efficient and effective visualization process. In other words, t-SNE allows for the visualization of complex data by projecting it onto a simpler, more understandable space. This technique is useful for identifying clusters or patterns in complex datasets, such as those found in hyperspectral imaging [163].

2. **Pre-processing techniques for medical images:** After collecting the raw medical image data, pre-processing techniques will be used. The pre-processing techniques for medical images are normalization, noise and band reduction, calibration, and spectral correction.
  - a. **Image Calibration (IC):** Dark and white reference images are obtained when the images are captured in the operation theatre. White references are acquired when the camera light is on. And the dark references are acquired by possession of the shutter of the camera. The calibrated image is denoted as  $I_C$ ,

$$I_C = 100 \times \frac{RI - DR}{WR - DR} \quad (1)$$

The raw input image, referred to as RI, is accompanied by two reference images - the white reference image, denoted as WR, and the dark reference image, represented by DR [164].

- b. **Noise and Band Reduction:** As soon as the image is calibrated, the band reduction is carried out to eliminate the extremely noisy bands that are useless owing to the sensor's poor performance. It will remove the bands from lower to higher bands [165].
- c. **Spectral Correction:** Spectral correction is used to correct the spectral noise from the image. A spectral-corrected picture has been produced using a correction matrix to multiply the signal.

$$I_{SC} = I_C \times SCM \quad (2)$$

Here,  $I_C$  is calibrated hyperspectral cube's single band. The correction matrix is denoted by SCM, where an individual band is a group of virtual band spectral correction constants [166].

- d. **Data Normalization:** It reduces the brightness levels caused by the light illumination. Normalization coefficients remain the spectral signatures' RMS value (Root Mean Square).

$$c[i, j] = \sqrt{\frac{\sum_{k=1}^B (I_C[i, j, k])^2}{B}} \quad (3)$$

$$I_{Norm}[i, j, k] = \frac{I_{SC}[i, j, k]}{C[i, j]} \quad (4)$$



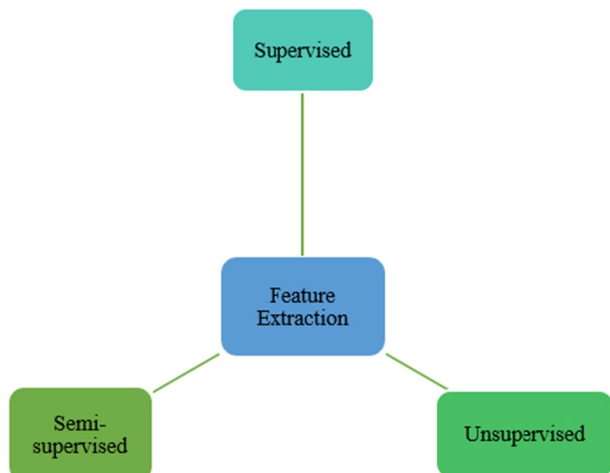
Here,  $I_{SC}$  is denoted as the spectral corrected cube with the dimensions  $R \times C \times B$  (Rows  $\times$  Columns  $\times$  Bands) [166].

## 5 Feature extraction

Extracting features is the method of recognizing and isolating significant features from data. Features are data bits that can be utilized to portray or differentiate between various information focuses. Feature extraction is crucial in many machine learning assignments, such as classification, regression, and clustering. Hyperspectral imaging employs feature extraction to extract features from the spectral signatures of pixels in an image. These features can then be utilized to recognize and classify materials in the image or identify environmental changes [167]. In the context of machine learning, the Hughes phenomenon refers to a situation where classification accuracy is severely reduced due to the high-dimensional properties of a few training samples. This phenomenon is a well-recognized challenge in the field of artificial intelligence and can cause significant problems in certain applications. Therefore, researchers and practitioners need to be aware of this issue and take appropriate measures to mitigate its impact.

Additionally, processing high-dimensional data eventually uses up computing resources and takes up space in data storage, according to the "curse of dimensionality" hypothesis [168]. Classifying hyperspectral images using Feature Extraction (FE) is challenging because of spectral unmixing, characterized by high intra-class inconsistency and inter-class resemblance. Nevertheless, Feature Extraction can help overcome these issues. Feature extraction methods categorized into three types are supervised FE method, unsupervised FE method, and semi-supervised FE method. These are used to extract the hyperspectral images' features [169]. Figure 17. represents the feature extraction methods.

**Fig. 17** Feature extraction techniques



1. **Unsupervised Feature Extraction:** these are used to extract the features from data without using labeled data. This is different from supervised feature extraction, which requires labeled data to extract features pertinent to a particular task. Unsupervised feature extraction is frequently used in hyperspectral imaging since it can be challenging or impractical to acquire labeled data for all the materials present in a hyperspectral image. Unsupervised feature extraction can extract features representing different materials in the image, even if the materials are unknown [168]. There are two primary types of unsupervised feature extraction techniques: Principal Component Analysis (PCA) and Independent Component Analysis (ICA). PCA determines the most significant variations in the data and projects the data in these directions. In contrast, ICA dissects the data into independent components that are as statistically independent as possible [170]. In the world of machine learning, three crucial techniques stand out: t-distributed Stochastic Neighbor Embedding (t-SNE), Autoencoder, and Generative Adversarial Networks (GANs). t-SNE allows for high-dimensional data to be condensed into a low-dimensional space while still maintaining the data's similar structure. Autoencoder is a specific type of neural network architecture that teaches itself to reconstruct input data from a lower-dimensional representation. Finally, Generative Adversarial Networks (GANs) are complex neural networks that comprise a generator, which creates new data similar to the training data, and a discriminator, which distinguishes between real and generated data [171].
2. **Supervised Feature Extraction:** The process of supervised feature extraction involves identifying pertinent features from labeled data. By training a machine learning model on this labeled data, it can then recognize those same features in unlabeled data. This technique is frequently utilized in various areas such as object recognition, image classification, and natural language processing [172]. There are several feature extraction techniques available, including Linear Discriminant Analysis (LDA) and Canonical Correlation Analysis (CCA). LDA is a powerful method that enables us to identify linear combinations of features that maximize the separation between different classes. Similarly, CCA is an effective technique that allows us to determine linear combinations of features that exhibit a high correlation between two sets of data [173]. Kernel PCA, Autoencoder, and Deep Boltzmann Machines are three techniques that can aid in the analysis of data. Kernel PCA is an extension of PCA that uses a kernel function to map input data into a higher-dimensional space, thus making it easier to distinguish between different types of data. In contrast, Autoencoder is a neural network architecture that learns to reconstruct input data from a lower-dimensional representation. Finally, Deep Boltzmann Machines are deep neural networks capable of learning intricate, layered representations of input data. These techniques have considerable potential for application in business and academic settings, providing a means to analyze complex data structures [174].
- a. **Nonparametric Weighted Feature Extraction (NWFE):** The NWFE method confidently employs weights to calculate weighted means for each sample. It also defines new nonparametric scatter matrices to generate more than L-1 features with utmost precision and accuracy. To define the nonparametric between-class scatter matrix for L classes, NWFE employs the following formula.

$$S_b^{NW} = \sum_{i=1}^L P_i \sum_{\substack{j=1 \\ j \neq i}}^L \sum_{l=1}^{N_i} \frac{\lambda_i^{(i,j)}}{n_i} \cdot (x_l^{(i)} - M_j(x_l^{(i)}))(x_l^{(i)} - M_j(x_l^{(i)}))^T \quad (5)$$

Here, the  $l^{th}$  sample from class I is indicated by  $x_l^{(i)}$ , the training sample size of class i is indicated by  $N_i$  and the prior probability of class i is indicated by  $P_i$  [175].



- b. **Linear Discriminate Analysis (LDA):** According to [176], the traditional LDA transforms original data into a discriminative subspace by minimizing intra-class scatter and maximizing inter-class scatter simultaneously through Fisher's ratio as a generalized Rayleigh quotient. Given a pairwise set is defined as  $\{(x_1, y_1), (x_2, y_2), \dots, (x_m, y_m)\}$ . The objective function for estimating the linear projection matrix P in multi-class LDA is given by:

$$\max_P \frac{\text{tr}(P^T S_b P)}{\text{tr}(P^T S_w P)} \tag{6}$$

Here, The variance or dispersion of data points between classes and within classes is defined by the  $S_b$  and  $S_w$  terms "between-class scatter matrix" and "within-class scatter matrix," respectively. By inserting the Lagrange multiplier  $\lambda$ , the optimization problem in Eq. (6) can be equivalently changed to one of the  $S_b P = \lambda S_w P$ , with the constraint of  $P^T S_w P = I$ . To solve the simplified optimization problem, one can use a technique called generalized eigenvalue decomposition (GED). The original Linear Discriminant Analysis (LDA) is highly susceptible to statistical degradation due to the complexity of high-dimensional noise caused by environmental and instrumental factors, as well as the limited availability of labeled samples. This issue is particularly severe when working with small-scale samples. The singularity of the two-scatter metrics  $S_b$  and  $S_w$ , which is prone to overfitting, is the main cause of degradation. The regularized LDA was presented with an extra l2-norm constraint on  $S_w$ , parameterized by  $\gamma$ , to improve generality and stability the below equation is useful

$$S_w^{reg} = S_w + \gamma I \tag{7}$$

By replacing  $S_w$  in Eq. (7) with the regularized  $S_w^{reg}$ , the solution can still be found using the GED solver in the regularized LDA.

3. **Local Binary Patterns (LBP):** According to [177], LBP, or Local Binary Patterns, is a powerful image analysis technique that measures image texture. It works by examining the surrounding area of each central pixel. To begin, LBP compares the intensity of the central pixel to the intensity of its neighboring pixels. If the central pixel's intensity is greater, it is assigned a value of 1; otherwise, it is assigned a value of 0. This generates a binary code that effectively summarizes the gray-level structure of the image. The LBP method picks particular nearby  $\{t_i\}_{i=0}^{P-1}$  and central pixel  $t_c$ , after taking into account a small circular neighborhood, denoted as P. The given equation computes the LBP

$$LBP_{P,R(t_c)} = \sum_{i=0}^{P-1} s(t_i - t_c) 2^i \tag{8}$$

$$s(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

In this case, P, R stands for the number of sample points and circle radius based on a neighbor set of central pixels  $t_c$ , that is circularly symmetric. In case  $x \geq 0$  and 0 otherwise, the neighboring pixels  $\{t_i\}_{i=0}^{P-1}$ , and  $s(x)$ .

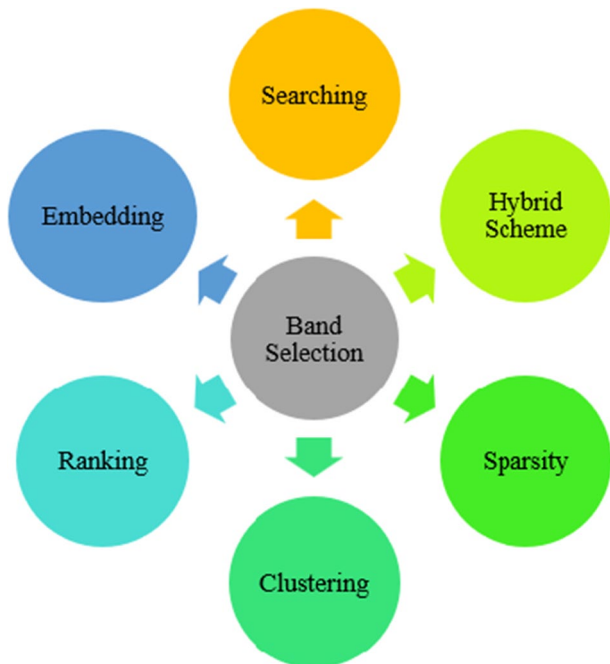
4. **Semi-supervised feature extraction:** A widely adopted approach in data feature extraction is the use of both labeled and unlabeled data, commonly referred to as semi-supervised feature extraction. This approach differs from supervised feature extraction

which solely relies on labeled data and unsupervised feature extraction which only uses unlabeled data. By leveraging a small set of labeled data along with a vast amount of unlabeled data, semi-supervised feature extraction can significantly boost the accuracy and efficiency of feature extraction, resulting in highly reliable and efficient outcomes [178]. Semi-supervised feature extraction is beneficial when labeled data is scarce or costly. This is often the case in hyperspectral imaging, where hyperspectral images can be large and complex, and labeling all the pixels in the image can be challenging and expensive. There are several noteworthy semi-supervised feature extraction methods, including Linear Local Tangent Space Alignment (LLTSA), Monogenic Binary Coding (MBC), Locality Preserving Projection (LPP), and Maximum Margin Projection (MMP) [179].

## 6 Feature (Band) selection

In the field of hyperspectral imagery, the selection of a suitable subset of bands for a specific purpose, such as classification, regression, or detection, is a critical task. These images often contain numerous bands, some of which may be redundant or noisy, making it imperative to carefully choose the appropriate bands to achieve accurate results [180]. The band selection process is a powerful tool that unlocks the potential of hyperspectral images, improving accuracy and efficiency in machine learning while reducing data dimensionality. According to [181], this article examines six main categories of current hyperspectral band selection methods: ranking-based, searching-based, clustering-based, sparsity-based, embedding-learning-based, and hybrid-scheme-based. Figure 18 represents the existing band selection techniques.

**Fig. 18** Band selection techniques



1. **Clustering-Based Methods:** Using a clustering-based method, the bands are confidently grouped into clusters, and the most representative bands from each cluster are selected to form the final subset. To accomplish this, the algorithm utilizes the highly reliable hierarchical clustering technique based on Ward's linkage. This groundbreaking paper serves as the first on HSI band clustering. To enhance cluster analysis, it is crucial to minimize the disparities within clusters while amplifying the disparities between them. To accomplish this, one can utilize information measures such as MI or Kullback-Leibler divergence to detect and eliminate repetitive bands. This process facilitates the precise identification of representative bands. Numerous other clustering-based band selection algorithms have been suggested in the literature, providing a range of options for improving the accuracy of the analysis. Except for Band Clustering [182], fuzzy clustering [183], and automatic band selection algorithms [184], most clustering methods derive from k-means, Affinity Propagation (AP), and graph clustering. These methods will be discussed in the following sections.
  - a. **K-means-based clustering methods:** According to [185], K-means is a popular and highly effective clustering technique. It partitions data into separate clusters based on similarity. Its simplicity, speed, and versatility have made it a preferred choice in data science, particularly for handling large datasets. It starts with a set of randomly chosen clusters, and then iteratively optimizes an objective function that measures the distance to a set of potential centers until the optimal cluster centers are identified. In particular, it seeks to minimize the objective function by dividing  $N$  bands into  $m$  clusters  $C$ , where  $C = \{c_1, c_2, \dots, c_m\}$  and  $c_j = (j_1, j_2, \dots, j_n)$  with  $1 \leq j_1 < j_2 < \dots < j_n \leq N$ .

$$\operatorname{argmin}_C \sum_{j=1}^m \sum_{b_i \in c_j} D(b_i, \mu_j) \quad (9)$$

The cluster center of  $C_j$  in the k-means algorithm is denoted by  $\mu_j$ , and the similarity metric  $D(\bullet, \bullet)$  computes the distance between a band and the cluster center to which it belongs.

- b. **Affinity Propagation-based clustering methods:** The exemplar-based AP clustering algorithm confidently overcomes the sensitivity of the k-means clustering algorithm to initial conditions and successfully identifies representative bands [186]. The AP algorithm expertly maximizes a function by taking into account both inter-band similarity and intra-band discriminative capability, which ultimately results in a flawless exemplar  $e$ .

$$H(e; \Theta) = \exp\left(\sum_{i=1}^N s(i, e_i) + \sum_{i=1}^N \log f_i(e)\right) \quad (10)$$

A similarity matrix  $s(i, e_i)$ , containing the eligibility of each band  $e_i$  to serve as the exemplar for the  $i^{\text{th}}$  bands are involved in Eq. (10). A coherence constraint, represented by the second term,  $\sum_{i=1}^N \log f_i(e)$ , indicates that a band must be its exemplar if it is selected as an exemplar by other bands.

- c. **Graph-based clustering methods:** According to [187], band selection can be formulated as a graph problem in graph theory. Within the HSI bands graph, each band is represented as a node. The edges linking these nodes denote the level of similarity between them. Utilizing a clustering technique, an affinity matrix  $A$  is created from this band similarity data. This matrix enables the graph to be clustered into subgraphs and subsequently identify the most representative bands. Band affinities

are represented by the affinity matrix  $A$ , where  $\sigma$  is a scaling factor and each affinity between pairs of bands is calculated as  $a_{i,j} = \exp(-f_i - f_j/2\sigma^2)$ . Graph-based clustering techniques, such as spectral clustering, are commonly used for clustering on the stacking eigenvectors of the affinity matrix  $\Gamma$ , which are defined as follows:

$$L = \Lambda^{-1/2}\Gamma\Lambda^{-1/2} \tag{11}$$

The affinity matrix's stacking eigenvectors are utilized for clustering in spectral clustering. The diagonal matrix  $\Lambda$ , which is calculated by summing over the rows of  $A$ , and the normalized graph Laplacian  $L$  are used to normalize the affinity matrix.

2. **Embedding learning-based methods:** These methods optimize application models such as classification, target detection, and spectral unmixing by combining them with band selection.

a. **Classifier learning-based methods:** The SVM classifier is an established and widely used method for hyperspectral image analysis due to its low sensitivity to imbalanced training samples. The RFE-SVM model, which stands for Recursive-Feature Elimination Support Vector Machine, is a highly effective approach for selecting hyperspectral bands and improving overall performance [188]. Weight values are obtained during the SVM classifier's training phase and are used as ranking criteria to remove unnecessary bands and improve the classifier. Recursive feature elimination, or RFE, aims to reduce generalization error by eliminating features that increase the margin. The predictive ability assessment,  $s^{REF}$  is computed as follows and is inversely proportional to the margin:

$$s^{REF} = \sum_{i=1}^D \sum_{j=1}^D \alpha_i \alpha_j \gamma_i \gamma_j \Phi(x_i, x_j) \tag{12}$$

Here, the  $i^{th}$  training samples and class labels are denoted by  $x_i$  and  $x_i \in \{-1, 1\}$  respectively. The kernel function used in SVM is  $\Phi(x_i, x_j)$ .

b. **Other learning-based methods:** It is worth considering the possibility of integrating band selection models into learning models to improve target detection and endmember extraction. In reference [189], to promote sparsity, a band sparsity term was incorporated into the objective function. This was accomplished through the incorporation of a sparsity-promoting prior, which was integrated into the iterative-constrained endmember algorithm. By introducing this term, it was possible to expand the sparsity-promoting capabilities of the algorithm, thereby improving its overall performance. This approach is particularly effective in a variety of applications, including signal processing, image analysis, and machine learning. As such, it represents a powerful tool for researchers and practitioners alike who are working to develop more effective and efficient algorithms for solving complex problems.

$$J = \eta \frac{RSS_B}{N} + \beta SSD_B + SPT + BST \tag{13}$$

Here,  $RSS_B$  is the residual sum of squares based on the convex geometry model,  $SSD_B$  is the term that describes the sum of squared distances, SPT represents the band sparsity-promoting term, while BST accounts for the weighted sum of band weights. The regularization parameters,  $\eta$ , and  $\beta$ , balance  $RSS_B$  and  $SSD_B$  in the objective function.

3. **Ranking-based methods:** Ranking-based methods are utilized to prioritize spectral bands based on a predefined criterion. These methods aim to determine the significance of each spectral band and select the top-ranked bands in a sorted sequence. By doing so, these methods provide a systematic approach to identifying the spectral bands that are most important for a given application. These methods can be divided into two types: unsupervised and supervised, based on whether labeled training samples are used.

A. **Unsupervised:** Unsupervised ranking-based band selection is a method to select bands from hyperspectral images without labeled data. It contrasts supervised band selection methods that require labeled data to train a machine learning model to select the most informative bands [190]. Unsupervised criteria consider the information, dissimilarity, or correlation of bands. Metrics such as variance, first spectral derivative, spectral ratio, contrast measurement, signal-to-Noise Ratio (SNR), third-order statistics (skewness), fourth-order statistics (kurtosis), kth order statistics, negentropy, entropy, and information divergence, are often employed to prioritize bands [191].

- a. **High-information criteria:** Specifically labeled bands must have a considerable information volume. Dissimilar classical information metrics, including information divergence and entropy, are utilized for grading to choose the bands [192]. The band-decorrelation approach confidently utilizes Kull-back-Leibler (KL) divergence to effectively remove any unnecessary or significant bands. The lower priority band is eliminated if the two bands' divergence value exceeds the threshold. In the future, mutual information remains to measure band dissimilarity. The covariance-based technique arranged all spectral bands via a matching filter and an adaptive coherence estimator to lessen their impact on target detection [193].
- **Information Entropy (IE):** Shannon entropy is a precise and widely accepted measure of the information content of a discrete random variable B, accurately defined by information theory. It is calculated based on the probability distribution  $p(b)$  of the variable, making it an essential tool to evaluate the amount of information conveyed by the variable.

$$H(B) = - \sum_{b \in B} p(b) \log p(b) \quad (14)$$

Subject to

$$\sum_{b \in B} p(b) = 1$$

$$p(b) = \frac{h(b)}{M \times N} \quad (15)$$

where B is a band,  $h(b)$  is its gray-level histogram, and  $M \times N$  is the total number of pixels in B.

- **First Spectral Derivative (FSD):** It regulates the bandwidth variable as a function of supplementary data

$$D_1 = \frac{\partial I(x, \lambda)}{\partial \lambda} \quad (16)$$

- **Second Spectral Derivative:** In hyperspectral images, SSD is used to explore the bandwidth variable as a purpose of additional data

$$D_2 = \frac{\partial^2 I(x, \lambda)}{\partial \lambda^2} \quad (17)$$

- b. **Low-Correlation Criteria:** If the specifically labeled bands consume lower mutual correlations, the aim is to reduce the selected bands' correlation; therefore, band selection is combined into an objective identification framework. Band representation is considered the required target signature; the remaining bands are unknown signature vectors. The CBS (Constrained Band Selection) approach uses Constrained Energy Minimization (CEM) to restrain a band depiction during the Band Correlation (BC) reduction [194].
  - c. **Large-Dissimilarity Criteria:** It is anticipated that the selected bands are different. A ranking-based example component analysis is suggested to quickly locate cluster centers across all bands [200]. To use this approach, it is not necessary to parameterize a probability density function. Rather, it is simply desirable to quantify the distances among all corresponding bands.
- B. Supervised ranking-based methods:** In the field of hyperspectral imaging (HSI), two methods that are commonly used to prioritize bands are supervised and unsupervised ranking-based methods. However, supervised methods are distinct from unsupervised ranking-based methods in that they rely on prior knowledge of HSI data to construct a band-prioritization criterion that can closely correlate with specific applications such as classification and spectral unmixing. This correlation is achieved through the use of labeled training data to train a model that can then be used to predict the importance of each band for a given task. By contrast, unsupervised ranking-based methods do not require labeled training data and instead use statistical methods to identify bands that are most relevant to a given task. While both methods have their advantages and disadvantages, the use of supervised methods can lead to more accurate and reliable results in certain applications. Supervised ranking-based approaches are separated into two types: spectral unmixing and classification criteria. Spectral unmixing-aimed-criteria: For spectral unmixing, it is employed. Orthogonal Subspace Projection (OSP), which depends on linear mixture models, is used for subspace projection to minimize subconscious noise and undesired signatures [191]. Classification-aimed criteria: Choosing the bands are used to assure optimum classification performance. An MMCA method (Minimum Misclassification Canonical Analysis) is designed to order bands through classification. Misclassification bands' error rate can be reduced by addressing the eigenvalue problem, and for this purpose, MMCA is used [195].
4. **Searching-based methods:** By converting band selection into an optimization problem through searching-based methods, we can confidently determine the best bands for creating an optimal solution based on a given criterion function. Two crucial issues arise in searching-based methods: 1) the criterion function and 2) the searching strategy. The criterion function can be similarity-based measurements such as Euclidean Distance (ED), Bhattacharyya distance, Jeffries–Matusita (JM) distance [196], Spectral Angle Mapping (SAM) [197], structural similarity index measurement [198], or information-based measurements such as Spectral Information Divergence (SID), transformed divergence, MI [199], and spatial entropy-based MI [200]. The search strategy determines

the best way to find an optimal or suboptimal solution. Based on the adopted searching strategy, searching-based methods can be grouped into three categories: incremental searching, updated searching, and eliminating searching.

- A. **Incremental searching:** To avoid the computationally prohibitive task of exhaustively testing all band combinations, incremental searching-based methods sequentially add new bands that optimize the criterion into current band subsets until a desired number of bands is selected. The Sequential Forwarding Selection (SFS) strategy is often implemented. These methods can be unsupervised or supervised, depending on whether labeled training samples are needed during the search process [201].
- a. **Unsupervised searching:** There are unsupervised searching-based methods that can iteratively add informative bands to enhance the accuracy of HSI data without requiring any prior knowledge. In [202], A new algorithm has been proposed to identify bands that exhibit significant skewness or kurtosis values. The algorithm is designed to detect these features with higher sensitivity and accuracy than existing methods. The proposed approach employs advanced statistical techniques to capture the complex distributions of data more effectively. In [203], To achieve acceptable performance in target detection and classification, we utilized Linear Prediction (LP) and OSP together to confidently assess the similarity between single and multiple bands. In [204], the utilization of spectral rhythm was found to be effective in enhancing the intermediary representation of Hyperspectral Imagery (HSI) data. Through iterative selection based on bipartite graph matching, the algorithm was able to identify the most informative and dissimilar bands. Additionally, the algorithm utilized convex set geometry to search for new vertices iteratively that maximize the largest simplex in the pixel space. As a result, the selected bands corresponding to these vertices had low correlations with each other. Overall, this approach has shown promising results in improving the representation of HSI data [205].
  - b. **Supervised searching:** Incorporating prior knowledge of hyperspectral imaging (HSI) data is a key strategy for improving class separability. Incremental searching-based methods represent a promising approach for achieving this objective. By leveraging prior knowledge, these methods can steadily refine the separability of classes over time. As such, they offer a powerful tool for enhancing the accuracy and effectiveness of HSI analysis in a variety of business and academic settings. In [206], the Band Add-On (BAO) algorithm was expertly developed and presented as a powerful tool. It utilizes the exact decomposition of SAM, enabling it to iteratively select the most effective bands. To optimize performance, it is necessary to increase the angular separation of two spectra in a spectral library. To effectively distinguish between two types of spectra, the BAO method was augmented with two band selection techniques: the average distance and Minimum Distance Methods (MDMs). By meticulously selecting the appropriate bands, the angular separation between the two categories was significantly enhanced, resulting in improved accuracy. Similarly, in [207], the Minimum Estimation Abundance Covariance (MEAC) MEAC algorithm incrementally selects dissimilar spectral bands to preserve classification information by minimizing the trace of the abundance covariance matrix using class spectral signatures



$$\operatorname{argmin}_{M^s} \left\{ \operatorname{trace}[(\hat{S}^T \hat{S})^{-1}] \right\} \quad (18)$$

In the above equation,  $M^s$  represents the selected band subset and the matrix  $S$  contains the spectral signatures of the classes in the selected band subset. Through the utilization of efficient SFS searching, the nonlinear parsimonious feature selection algorithm has been able to effectively and iteratively maximize the classification rate estimate from the Gaussian mixture model classifier in carefully selected spectral bands.

- B. Updated searching-based methods:** The latest searching-based methods confidently optimize the predefined evaluation criterion by iteratively replacing elements of the current band subset with new ones as required during the searching procedure. Aside from simple searching strategies such as the sequential forward-floating search [208], the branch-and-bound search [209], the steep ascent search, and the constrained search [210], evolutionary algorithms have been adopted for band searching, such as Particle Swarm Optimization (PSO) [208], Adaptive Simulated Annealing [211], Genetic Algorithms (GAs) [212], Firefly Algorithms (FAs) [213], Differential Evolutionary Algorithms [214], and Ant Colony Optimization [215]. Similar to incremental searching-based methods, similarity or information measurements can be used as an objective function.
- a. **Classifier Independent Methods:** The latest searching-based approaches incorporate an objective function that evaluates class separability based on specific metrics while disregarding any classification accuracy from a true-classifier standpoint. In [216], The Clonal-Selection Feature-Selection algorithm can meticulously select a subset of bands that optimizes the averaged JM distance among various classes. This algorithm employs a rigorous approach to ensure that the selected subset is a true representation of the entire dataset. By selecting the optimal subset, the algorithm facilitates the identification of the most significant and relevant features, thus enhancing the accuracy of the classification process.
  - b. **Classification-dependent methods:** Through extensive evaluation, the authors in [217] determined that utilizing the accuracy of a genuine classifier in certain updated searching-based techniques can lead to a highly effective objective function. Our findings indicate that nature-inspired algorithms such as Gravitational Search, Harmony Search, PSO, FA, and Bat algorithms are the most optimal choices for selecting bands that maximize the accuracy of the Optimum Path Forest classifier.
- C. Eliminating searching-based methods:** The "eliminating search-based" approach to select the best bands for a task. This method starts with all the bands and removes the unnecessary ones until we reach the desired number of selected bands. Sequential Backward Selection (SBS) is a common way to do this [218].
5. **Sparsity-based methods:** According to the sparsity theory, each band can be accurately and efficiently represented through sparse usage of nonzero coefficients associated with atoms in a suitable basis or dictionary. Sparsity-based band selection methods use sparse representation or regression to reveal specific underlying structures within HSI data. To



find representative bands, an optimization problem with sparsity constraints is solved [219]. Also, the current sparsity-constrained methods are categorized into Sparse Non-negative Matrix Factorization (SNMF)-based, sparse representation-based, and sparse regression-based approaches.

- A. **Sparse nonnegative matrix factorization-based methods:** According to [220], Spectral Non-negative Matrix Factorization (SNMF) breaks down a data matrix in hyperspectral imaging into a set of bases and encodings. The basis matrix in SNMF is non-negative, while the coefficient matrix is negative and sparse. The non-negative constraint in both matrices is responsible for the parts-based feature of SNMF. This is because only additive combinations are allowed. SNMF was initially designed to solve the HSI band selection problem, which involves selecting representative bands by clustering the sparse coefficients. The technique aims to factorize the HSI band matrix  $B$  into two unknown matrices: the dictionary matrix  $W \in R^{D \times r}$  and the sparse coefficient matrix  $H \in R^{r \times N}$ . This is achieved by optimizing the objective function, we will get the below equation;

$$\min_{W, H} f_r = \frac{1}{2} B - WH_F^2 \quad (19)$$

Subject to  $W, H \geq 0$ , and  $\|h_i\|_0 \ll r, 1 \leq i \leq N$

To achieve a desired low-rank  $r$ , the subscript  $r$  in  $f_r$  is used. Each column of  $H$  contains a cluster or subspace of each band it belongs to, with the most significant entry in each column representing it. The constraint  $\|h_i\|_0$  means that each column vector  $h_i$  is sparse, and the number of nonzero entries is significantly smaller than the dimensionality  $r$ . Unfortunately, the ED distance measurement in Eq. (19) inaccurately represents the error between  $X$  and its approximation  $H$ , as the Gaussian distribution assumption behind the ED measurement contradicts the nature of HSI data.

- B. **Sparse representation-based methods:** Sparse representation-based methods employ either manual definition or learning of a dictionary in advance to select informative bands based on sparse coefficients. In these methods, the dictionary is a set of basis functions that represent the input signal. By selecting a sparse representation of the signal, that is, a representation that uses only a small number of basis functions, the method can identify the most informative bands in the signal. An algorithm was proposed for selecting bands from hyperspectral images, which is based on sparse representation. In this algorithm, the hyperspectral image bands were sparsely represented using a dictionary learned by K-SVD. The algorithm then ranked the bands based on their sparse coefficients, using majority voting. Finally, the bands with high occurrences in the histograms of sparse coefficients were selected [221].
- C. **Sparse regression-based methods:** According to [222], Sparse regression-based methods can be used to solve the band selection problem. To do this, the problem is transformed into a sparse linear regression problem, which uses training samples and their class labels. The sparse coefficients obtained from the best solution are then used to select the bands that provide better class separability. To ensure sparsity, a constraint is imposed on the linear regression between the training samples and their class labels. The Least Shrinkage And Selection Operator (LASSO) is used to obtain the solution.

$$\operatorname{argmin}_{W_c} \|y_c - B^T W_c\|_F^2 + \lambda \|W_c\|_1 \quad (20)$$

The given equation consists of class labels, where 1 indicates belonging to the specific class, and 0 denotes the opposite. The parameter  $\lambda$  controls the  $L_1$  penalty term's contribution. The bands were chosen based on the ranking of their coefficients for all classes. A combined framework for band selection was developed by coupling the LASSO model and a new separability measure that employs the Hilbert–Schmidt independence criterion.

6. **Hybrid scheme-based methods:** Hybrid schemes are a powerful tool in identifying the most appropriate bands, allowing for versatility and efficiency. These methods often combine clustering and ranking schemes to create hybrid algorithms. In a study referenced as [223], the Spectral Separability Index (SSI) algorithm integrates clustering, ranking, and searching schemes to select the most optimal band combination for classification. First, the original bands are grouped into clusters based on their spectral Correlation Coefficients (CCs). Then, a representative band from each cluster is chosen by ranking the entropy of the bands in the same cluster. Finally, the best band combination is selected by maximizing the SSI separability.

## 7 Hyperspectral image classification

The process of hyperspectral image classification is a highly accurate and effective method. It involves assigning pixels in an image to specific classes based on their spectral signatures. Hyperspectral images contain a wealth of information about object reflectance, with hundreds or even thousands of narrow, contiguous wavelength bands. This enables us to perform a range of applications, including precise mineral mapping, vegetation analysis, and urban land-use mapping [224]. The primary objective of hyperspectral image classification is to accurately identify and categorize different objects within an image based on their spectral properties. This is accomplished by extracting features from the spectral signatures of the pixels, which are then utilized to train a machine learning model. With the help of this model, we can confidently classify the pixels, ensuring accurate object identification and classification. The objects present in the image can be anything from soil, vegetation, water, buildings, or any other type of object that possesses a unique spectral signature [225]. We have divided the classification techniques into traditional machine learning neural networks in hyperspectral image classification. In traditional machine learning, we have divided into supervised, semi-supervised, and unsupervised classification techniques. Coming to the neural network classification model, we have subdivided it into traditional neural networks and deep learning (Fig 19).

### 7.1 Traditional machine learning

#### 7.1.1 Supervised machine learning

According to [116], Supervised machine learning classification is an algorithmic paradigm that facilitates the categorization of data into distinct groups. This is achieved by training the algorithm on a labeled dataset, where each data point is assigned a known label. The algorithm then identifies patterns in the data associated with each label, which

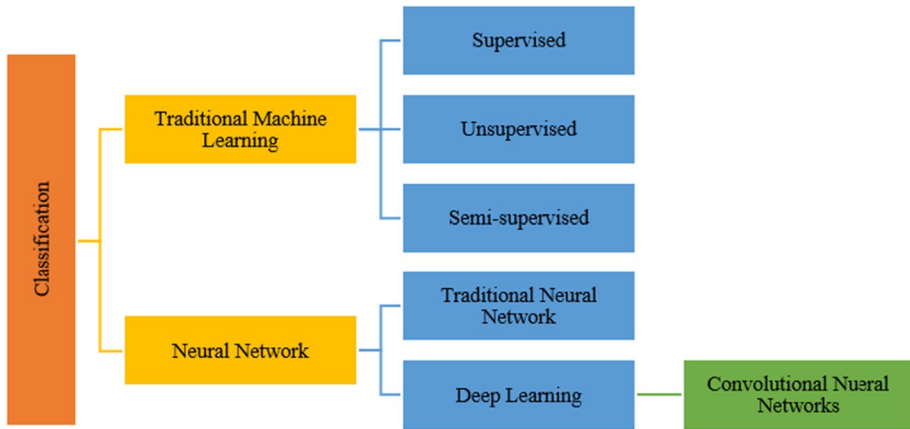


Fig. 19 Hyperspectral Image Classification

can subsequently be utilized to classify new data points that have not been previously encountered. Standard supervised learning algorithms include Random Forest (RF), Logistic Regression (LR), Artificial Neural Networks (ANN), Decision Trees, Support Vector Machine (SVM), Gaussian Naive Bayes, and Nearest Neighbours. These algorithms are widely employed in business and academic settings, owing to their efficacy in predicting and classifying outcomes.

**A. Support Vector Machines (SVM):** SVMs are widely recognized as powerful machine learning algorithms utilized in various applications such as classification and regression analysis. SVMs derive their strength from the concept of decision planes, which effectively determine the boundaries between objects with different class memberships. The SVM algorithm is a robust and reliable tool that efficiently categorizes data into distinct training and testing sets. It skillfully identifies the hyperplane that optimizes the distinction between classes, with minimal room for error. SVMs have demonstrated their efficacy across a range of applications, from image and text classification to bioinformatics, and have a well-established record of success. In the training set, each instance is assigned a target value, making SVMs an excellent choice for solving complex classification and regression problems with high accuracy. This target value represents the class the instance belongs to [226]. According to [227], the SVM algorithm then uses this information to learn each class's characteristics and create a model that can be used to classify new instances. The objective of this model is to identify the optimal hyperplane that effectively separates the distinct classes within the training data. Once trained, it can classify new, unseen data accordingly.  $X$  denotes the input dataset, while  $Y$  denotes the output dataset. Through the utilization of these datasets, this model can generate precise and reliable predictions. The training set will be defined as  $\{(x_1, y_1), (x_2, y_2), \dots, (x_m, y_m)\}$

$$Y = f(x, \alpha) \quad (21)$$

Here, Kernel function parameters  $\alpha$  are fine-tuned for accurate classification by the SVM classifier, which transforms data into a higher-dimensional space for easier class separation. Various types of kernel functions can be used with SVM, including poly-

mial, linear, and radial basis functions. Each kernel function has unique characteristics that significantly impact SVM classifier performance. Polynomial kernels are helpful for data that has non-linear relationships, while linear kernels are used for linearly separable data. Radial basis functions are often used when the data has no clear separation between classes. The choice of kernel function is an important consideration when using SVM, and it is often determined through experimentation and testing. The SVM decision function is given by

$$f(x) = \sum_{i \in s} \alpha_i y_i K(x_i, x_j) + b \tag{22}$$

In this case, the kernel function  $K(x_i, x_j)$  reflects the machine learning mathematical technique that converts data into a higher-dimensional space. The subgroups of the training sample are indicated by  $s$ .

Several kernels are commonly used:

Linear:

$$K(x_i, x_j) = (x_i, x_j) \tag{23}$$

Polynomial:

$$K(x_i, x_j) = (\gamma(x_i, x_j) + c)^d \tag{24}$$

Radial basis function:

$$K(x_i, x_j) = \exp\left\{ \frac{\|x_i - x_j\|^2}{2\sigma^2} \right\} \tag{25}$$

Sigmoid:

$$K(x_i, x_j) = \tanh(\gamma(x_i, x_j) + c) \tag{26}$$

**B. Maximum Likelihood Classification:** According to [228], the Maximum Likelihood (ML) classification method is often preferred as it can yield better results in hyper-spectral remote sensing images. This is particularly true when the training samples are normally distributed, as the classification method obtained through ML tends to be more effective in such cases. In remote sensing, a ground feature image can use its spectral feature vector  $X$  to locate a corresponding feature point in the spectral feature space. Each feature point from a similar feature will form a cluster of certain probability in the feature space. The conditional probability  $P(\omega_i|X)$  of a feature point ( $X$ ) falling into a certain cluster ( $\omega_i$ ) can be used as a component category decision function, which is called a likelihood decision function. Assuming that  $g_i(x)$  is a discriminant function, the probability  $P(\omega_i|x)$  that a pixel  $x$  belongs to class  $\omega_i$  can be expressed as

$$g_i(x) = P(\omega_i|x) \tag{27}$$

So basically, there's this thing called the Bayesian formula that helps you figure out the probability of something happening based on new information that we will get.

$$g_i(x) = P(\omega_i|x) = P(\omega_i|x), P(\omega_i)/P(x) \tag{28}$$

In this case,  $P(\omega_i|x)$  is the conditional probability that  $x$  is a member of  $\omega_i$ ,  $P(\omega_i)$  is the prior probability, and  $P(x)$  is the probability that  $x$  is not a member of the category. One

approach to hyperspectral data classification is maximum likelihood, which assumes that the data has a normal distribution. Equation (29) provides the discriminant formula, which makes distinctions between various classes.

$$g_i(x) = P(\omega_i|x) = P(\omega_i) = \frac{P(\omega_i)}{(2\pi)^{k/2} |\sum_i|^{1/2}} \exp\left[-\frac{1}{2}(x - u_i)^T \sum_i^{-1} (x - u_i)\right] \quad (29)$$

In this case,  $i$  denotes the number of classes,  $k$  the number of features, covariance of the matrix of the  $i^{\text{th}}$  class is provided as  $\sum_i$ , determinant of the matrix  $\sum_i$  is given as  $|\sum_i|$ , and  $u_i$  is the mean vector.

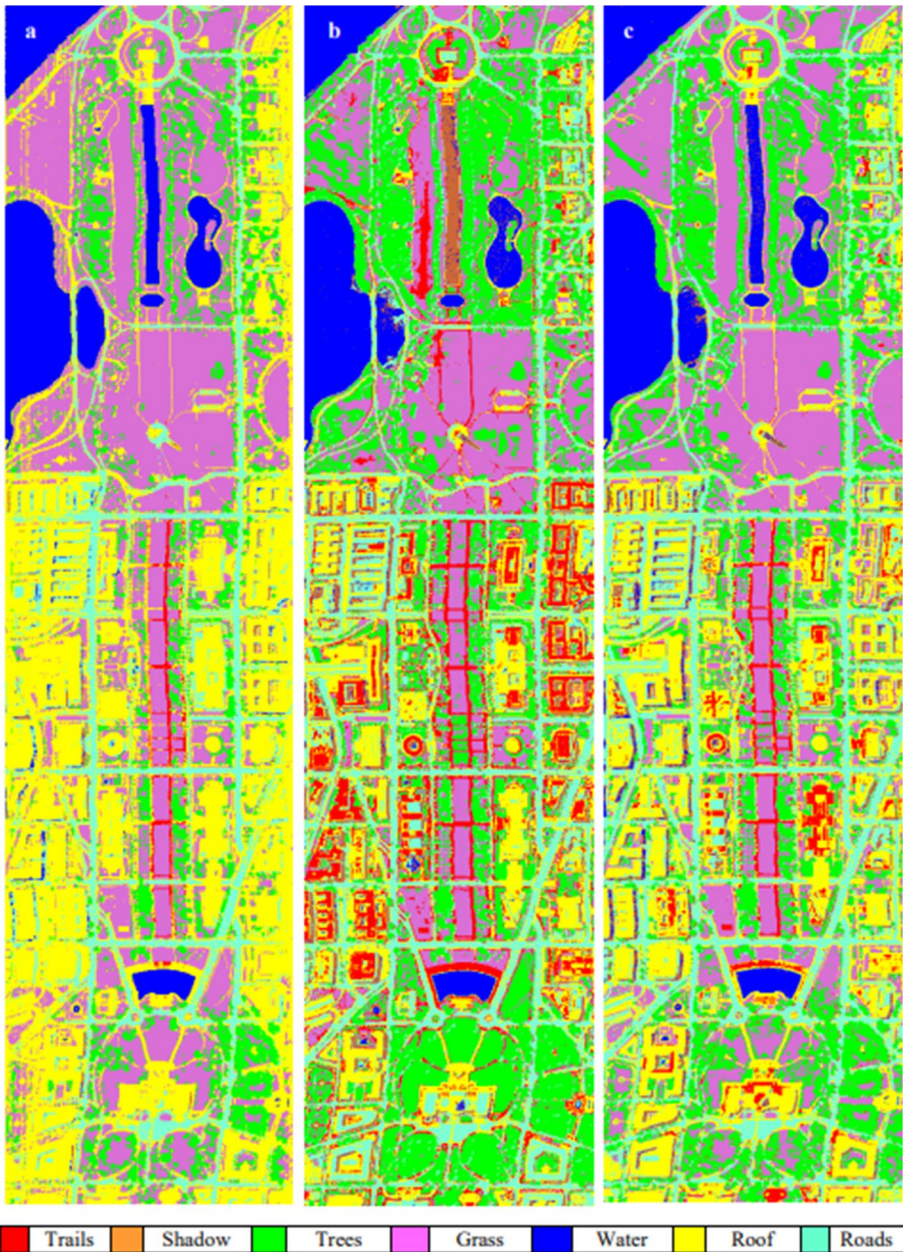
- C. **Spectral Angular Mapper (SAM):** According to [229], the Spectral Angle Mapper program computes the angle between two spectra and considers them as vectors with dimensions equivalent to the number of bands to establish spectral similarity. On the other hand, spectroscopy is a technique that determines the molecular structure by measuring the radiant intensity and energy of the interaction between light and the subject of interest. In absorption spectroscopy, the element that interacts with light is passive. It absorbs specific photons based on their wavelength, resulting in a spectral signature. The light not absorbed can either pass through the chemical sample or be diffusely reflected on it. Once the diffuse reflectance spectrum is obtained, it must be processed to identify, classify, or discriminate the elements. The SAM method generalizes this geometric interpretation to  $n$ -dimensional space. It uses the following equation to determine similarity:

$$a = \cos^{-1} \left( \frac{\sum_{i=1}^{nb} t_i r_i}{\sqrt{\sum_{i=1}^{nb} t_i^2} \sqrt{\sum_{i=1}^{nb} r_i^2}} \right) \quad (30)$$

Here, the number of bands is denoted by  $nb$ , the pixel spectrum is represented by  $t$ , and the reference spectrum is depicted as  $r$ . Figure 20 shows the classification results of Washington DC Mall hyperspectral image using ML, SAM and SVM.

- D. **Decision Tree (DT):** The construction of the decision tree is a confident and reliable process, achieved through the recursive division of training data into subsets based on attribute values until the stopping criterion is reached. The stopping criterion could be the maximum depth of the tree or the minimum number of samples required to split a node. This process of dividing the subsets based on the attribute values is continued until the stopping criterion is met. In the training phase, the Decision Tree algorithm selects the best attribute to divide the data based on a metric such as entropy or Gini impurity. The metric measures the level of impurity or randomness in the subsets of the data. The objective is to find the attribute that results in the maximum information gain or reduction in impurity after the split. This attribute selection process continues until the tree is fully grown [230].
- E. **Random Forest (RF):** Random Forest is an ensemble learning technique that combines multiple decision trees to improve accuracy in classification and regression tasks. It can handle large datasets and identify influential features to improve model interpretability. In a Random Forest, each decision tree is constructed using a random subset of the training data and a random subset of the features. This helps to reduce overfitting and improve the accuracy and generalization of the model. When making predictions, the Random Forest algorithm aggregates the predictions of all the individual decision trees to arrive at a final prediction. Overall, Random Forest is a powerful and flexible algorithm that can be used for various machine-learning tasks [231].





**Fig. 20** Classification results. **a** ML classification, **b** SAM classification, and **(c)** SVM classification [228]

- F. CART (Classification and Regression Trees):** According to [232], It is a rule-based data mining technique used for classification and regression tasks. CART has the learning procedure in two stages: 1. Selecting the tree structure and 2. Determining the predictions at the leaf node. It will isolate the input data into independent variables with

the best degree of purity. Where the same land use type is the source of the leaf nodes. At each node, there are numerous criteria for data partitioning, one of which is the Gini index, which may accommodate nominal values. The Gini index at node  $t$  is determined using.

$$Gini(t) = \sum_{i \neq j} P(w_i) \times P(w_j) \quad (31)$$

Here,  $P(w_i)$  is the comparative regularity of  $i^{\text{th}}$  class. The tree expansion procedure is frequent until the maximum spotlessness at the leaf nodes is obtained. If the decision tree models a target variable with nominal values, it is referred to as a classification tree; if it models a target variable with continuous values, it is referred to as a regression tree.

- G. **K-Nearest Neighbour (K-NN):** According to [233], the K-NN algorithm is a non-parametric method widely used for classification in pattern recognition. The main principle of K-NN is that the category of a data point is determined according to the classification of the nearest K neighbors. If we have a training set  $T = \{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)\}$ , here N is the Number of training entities. Here,  $x_i \in R^d$  denotes the feature vectors,  $y_i \in y = \{c_1, c_2, \dots, c_m\}$  depicts the classification labels. An input  $x$  is given as  $i = 1, 2, \dots, N$ , from this we can obtain the K-nearest neighbors  $N_K(x)$  by computing the distance with training entities.

**Distance Metrics Used in K-NN:** The KNN algorithm is used to identify the nearest groups or points to a query point. However, we need to use a metric to determine the closest groups or points for a query point. For this purpose, we use the following distance metrics:

- i. **Euclidean Distance:** The Euclidean distance measures the Cartesian distance between two points in a plane or hyperplane. It can be visualized as the length of a straight line that connects the two points in question. This metric is beneficial for calculating the net displacement between two states of an object.

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (32)$$

- ii. **Manhattan Distance:** The Euclidean distance is advantageous when we are interested in calculating the total distance traveled by an object rather than just its displacement. To compute this metric, we sum the absolute differences between the coordinates of the points in n-dimensions.

$$d(x, y) = \sum_{i=1}^n |x_i - y_i| \quad (33)$$

- iii. **Minkowski Distance:** It's worth noting that both the Euclidean and Manhattan distances are exceptional cases of the Minkowski distance.

$$d(x, y) = \left( \sum_{i=1}^n (x_i - y_i)^p \right)^{\frac{1}{p}} \quad (34)$$

**How to choose K value** In the k-nearest neighbors (k-NN) algorithm, the value of k is crucial as it determines the number of neighbors that will be considered. It's essential to choose an appropriate value of k based on the input data. For instance, if the input data

contains significant outliers or noise, a higher value of  $k$  may be more suitable. To avoid ties in classification, choosing an odd value for  $k$  is recommended. Cross-validation methods can help select the best  $k$  value for a given dataset.

**H. Multivariate Adaptive Regression Splines (MARS):** In [234], the authors have given that a nonparametric regression technique called Multivariate Adaptive Regression Spline (MARS) can spot interactions and nonlinear correlations between response and predictor variables. Additionally, the MARS technique can automatically choose crucial modeling variables. The equation below can be used to display the MARS model estimator.

$$f(x) = a_0 + \sum_{m=1}^M a_m \prod_{k=1}^{km} [S_{km} \cdot (x_{v(k,m)} - t_{km})] \tag{35}$$

Based on the least Generalised Cross-Validation (GCV) value, MARS uses stepwise forward and backward stepwise algorithms to determine knots (automatically) from the data. In other words, the chosen knot point has the lowest GCV value. The algorithm for determining knots uses the modified GCV formula as a criterion.

$$GCV(M) = \frac{(\frac{1}{N}) \sum_{i=1}^N [y_i - \hat{f}_m(x_i)]^2}{[1 - \frac{C(M)}{N}]^2} \tag{36}$$

Here,  $M$  is represented as a nonconstant basis function, the number of parameters in the model is  $C(M)$ , and the matrices bias function is  $B$ .

The significance of the parameters of the MARS model is tested in two stages - simultaneous testing and partial testing. The F test or Fisher test is used as the test statistic. The formula used for this test is as follows:

$$F = \frac{Sse/k}{\frac{Sse}{n} - k - 1} \tag{37}$$

**I. Minimum Distance Classification:** The Minimum Distance Classifier (MDC) technique is a highly reliable and widely accepted classification method that accurately categorizes pixels in the feature space based on their distance. It is a commonly accepted notion within the field of feature space that feature points of the same class tend to form clusters. These feature points determine the mean vector, which acts as the category's center. Additionally, the covariance matrix is precisely computed to describe the dispersion of the surrounding points. Points are then measured consistently and reliably for each category [228]. To identify whether the two modes are similar or not, a similarity measure is used. This measure confidently asserts that modes can be considered similar if their feature differences fall below a certain threshold. The technique creates decision-making regions by collecting different training sample points, and similarity is measured using distance as the primary metric. There are various methods for calculating distances, such as Ming's, Mahalanobis', absolute value, Euclidean, Chebyshev's, and Barth's distances. The effectiveness of Mahalanobis and Barth-Parametric distances in classification is widely acknowledged due to their ability to take into account the mean vector and



the distribution of each feature point around the class center. It's worth noting, however, that their computation requires more data than other distance criteria [235].

### 7.1.2 Unsupervised machine learning

This technique is used to group data points into clusters based on their similarities without using labeled data. This method differs from supervised machine learning classification, which relies on labeled data to train a model to classify new data points. Unsupervised machine learning classification is often used in exploratory data analysis to discover hidden patterns and relationships. For instance, it can be used to identify groups of customers with similar buying habits or groups of the population with different health risks [236]. Standard unsupervised machine learning algorithms are Principal Component Analysis (PCA), Adaptive Resonance (AR), Self-Organizing Maps (SOM), Artificial Neural Network (ANN), ISODATA, and Clustering. In clustering, we have K-means Clustering, Spatial Clustering, spectral clustering, fuzzy c-means clustering, mean shift clustering, Density-Based Clustering (DBSCAN), Balanced Iterative Reducing and Clustering using Hierarchies (BIRCH), Hierarchical Clustering

**A. Principal Component Analysis (PCA):** It is one of the most well-known and often employed dimensionality reduction techniques that use statistical measurements. To extract the information from the informative bands, PCA performs orthogonal transformations to transform HSI's highly correlated image bands into a set of linearly uncorrelated variables. According to [237], if  $x$  is the pixel vector of the hyperspectral image data, it can be represented as  $X_n = \{x_{n1}, x_{n2}, x_{n3}, \dots, \dots, x_{nF}\}^T$  with all pixel values  $X_1, X_2, X_3, \dots, \dots, X_S$  at one parallel pixel location of the hypercube or data matrix. The  $n$  represents the  $n^{th}$  number of pixels from  $s$ . here, the hypercube is represented by  $D$  of size  $F \times S$ , where  $S = X \times Y$ . To calculate mean vector,  $M$  of all image vectors:

$$M = \frac{1}{S} \sum_{n=1}^S X_n \tag{38}$$

The covariance matrix is calculated by using Eq. (40)

$$C = \frac{1}{s} II^T \tag{39}$$

Here,  $I$  is the zero-mean image as  $I = \{I_1, I_2, I_3, \dots, I_n\}$  produced from  $I_n = x_n - M = \{I_{n1}, I_{n2}, I_{n3}, \dots, I_{nF}\}^T$ . Now, covariance matrix, the variable  $C$  is computed to perform eigenvalue decomposition. This process follows a specific format which can be described as:

$$C = VEV^T \tag{40}$$

Here,  $V$  is the orthogonal matrix, where  $F$  is the dimension eigenvectors ( $V_1, V_2, V_3, \dots, V_F$ ) are used to create the orthogonal matrix, and  $E = \text{diagonal}(E_1, E_2, E_3, \dots, E_F)$  is a diagonal matrix composed of the corresponding eigenvalues ( $E_1, E_2, E_3, \dots, E_F$ ) The eigenvectors in this case are referred to as principal components (PCs). Currently, a new feature subspace  $w$ , a  $F \times k$  dimensional matrix with  $k \leq F$  and frequently  $k \ll F$ , is created by selecting  $k$  eigenvectors. Divergence analysis, discriminant analysis, and other methods can be used to select  $k$  eigenvectors,

such as sorting the eigenvectors in descending order and selecting the top  $k$  principal components. Ultimately, the modified PCA pixel vector  $Y$  can be acquired as follows:

$$Y = w^T \times I \quad (41)$$

Here, the original image data is represented as  $d$ , which can be created as  $d = (w * Y) + M$

- B. Self-Organising Maps (SOM):** This is an artificial neural network that takes high-dimensional data and maps it to a lower-dimensional space. The resulting map comprises a set of nodes that preserve the distribution of the input data while also representing the entire dataset. The lattice structure can be of any shape, but a finite two-dimensional rectangular grid is the most commonly used. Each node in the SOM is associated with a vector  $z$  that has the same dimension as the input data points [238]. The SOM operation consists of two main parts: training and labeling. Although they can be combined into a single online process, we will describe them separately. In both parts, the goal is to identify the best matching unit (BMU) for a given input vector. The BMU is the node closest to the input vector under a specified distance metric, typically Euclidean distance. The best matching unit is given as

$$BMU(x_i) = \operatorname{argmin} k(x_i, Z) \quad (42)$$

Here, input data is given as  $X = [x_1, x_2, x_3, \dots, x_n]$ ,  $x_i \in R^d$ ,  $m \times m$  grid of the SOM nodes a given as  $Z = [Z_{1,1}, Z_{1,2}, Z_{1,3}, \dots, Z_{m,m}]$ ,  $Z_i \in R^d$ , and  $k$  is the distance metric. During the Self-Organizing Map (SOM) training phase, the BMU (Best Matching Unit) is identified, and its neighborhood function is calculated. The SOM adjusts neighboring nodes towards the input vector and labels the input vector with the BMU's SOM coordinates during the labeling phase. After training the map, if two points are located near each other in the input data space, then they will also be mapped to nodes that are positioned close to each other on the SOM grid. In other words, the SOM grid maintains the spatial relationships between various data points while preserving the topology of the input space. Gradient descent is analogous to the Self-Organizing Map (SOM) training procedure. However because the original SOM lacked an objective function, it's not the same. The index on the SOM of BMU for an input vector  $x_i$  from Eq. (43) can be represented as  $u_i^* = BMU(x_i)$ . For every SOM node  $Z_j$ , the updated steps are provided as

$$z_j^{s+1} = z_j^s + \alpha \beta(t, u_j, u_i^*) k(x_i, z_j^s) \quad (43)$$

where  $\alpha > 0$  is the learning rate for iteration, the neighborhood function is represented as  $\beta(t, u_j, u_i^*) > 0$ , each training epoch is denoted as  $t$ , and the iteration within the epoch is denoted as  $s$ . The learning rate is denoted by  $\alpha$ , which determines the speed at which the map adjusts to the input data. The neighborhood function  $\beta(t, u_j, u_i^*)$  Preserves the topology by training nodes that are spatially far from  $u_i^*$  with a lower magnitude. During training, the update radius ( $t$ ) is used to characterize the learning process. As the training progresses, this radius decreases to enable the model to learn the rough distribution topology before fine-tuning the local areas of distribution [239].

- C. Active Learning:** It is a machine learning method that selects the most informative samples from an unlabeled dataset and uses human input to label them, thereby training a supervised machine learning model. This approach aims to reduce the dependence on a large labeled training dataset. Two primary types of Active Learning

(AL) are stream-based and pool-based. In stream-based AL, the algorithm receives each unlabeled sample one at a time and decides whether to request a label. In contrast, pool-based AL involves a large pool of unlabeled samples presented to an AL acquisition function for selection and manual labeling. In an Active Learning framework, a supervised machine learning algorithm and an acquisition function play crucial roles [240].

- D. **ISODATA (Iterative Self-Organizing Data Analysis Technique):** The ISODATA algorithm is a commonly used unsupervised classification method that extends the K-Means algorithm. It selects the number of clusters automatically using heuristics. According to [241], the ISODATA algorithm assumes that each class follows a multi-variate normal distribution and requires each class's means and covariance matrices. It follows an iterative process where arbitrary cluster centers are assigned initially, and the cluster means and covariance are calculated. Then, each pixel is classified to the nearest cluster. New cluster means, and covariance are calculated based on all the pixels in each cluster. This procedure is iterated until the change between iterations is considered "low enough." The modification can be quantified by measuring the distances the cluster means have changed from one iteration to the next or by the percentage of pixels that have changed between iterations.

In more detail, the steps in ISODATA clustering are as follows:

- i. Specify the number of clusters.
- ii. The clustering algorithm will then proceed to select the initial cluster centers and assign the pixels to them accordingly.

$$x \in i \text{ if } |\omega(x) - \omega_i| < |\omega(x) - \omega_j| \text{ for all } j \neq i \quad (44)$$

Here, the cluster centers for cluster  $i$  and  $j$  are given as  $\omega_i$ , and  $\omega_j$ , and  $x$  is the position of the feature vector.

- iii. To calculate the new class mean, we confidently compute the average of the pixel values that are assigned to the class, which serves as the definitive center for class  $i$ .

$$\omega_i = \frac{1}{Q_i} \sum_{x \in i} \omega(x), i = 1, 2, 3, \dots, K \quad (45)$$

Here,  $K$  denotes the number of clusters,  $Q_i$  is the number of pixels in class  $i$ , and the cluster covariance is also calculated at the same time.

- iv. So basically, the pixels are assigned to the closest cluster.
- v. The determination to calculate the means and covariance of the new cluster is truly inspiring.
- vi. Repeat steps 4 through 5 if the difference between the initial and new clusters is not minimal enough. If not, the clustering process is over.

- E. **Clustering:** Clustering categorizes data points into distinct groups or clusters based on their similarities and differences. The objective of clustering is to group data points that are alike and separate those that are dissimilar. Clustering is a technique for organizing objects to make them easier to understand and analyze [242]. The types of clustering algorithms have been listed below.

- **K-means Clustering:** K-means is a widely used unsupervised learning algorithm that easily groups data into a specified number of clusters. If a set of observations are depicted as  $X = (x_1, x_2, \dots, x_n)$ , where each observation is  $x_i \in R^d$ , i.e.,  $x_i = [x_{i_1}, x_{i_2}, \dots, x_{i_d}]$ , here,  $d$  is represented as several spectral channels. To group each observation cluster into several clusters,  $k$  is fixed a priori value, i.e.,  $k < n$ . K-means will calculate the centers of the  $k$  groups by optimizing the error of each group as

$$\min \sum_{j=1}^k \sum_{i=1}^{n_k} \|x_i^j - c_j\|^2 \quad (46)$$

Here, the Euclidian distance between a data point  $x_i^j$  of the cluster,  $j$  is depicted as  $\|x_i^j - c_j\|^2$ , the cluster center is given as  $c_j$ , and the observations within each cluster are given as  $n_k$ . The K-means algorithm effectively extracts valuable insights from a dataset, especially when identifying the most suitable distance metric. However, the results can vary significantly depending on slight parameter changes and initial center selection. Therefore, it is crucial to initialize the process correctly to ensure the final output is the best solution [243].

- **Hierarchical clustering:** Hierarchical clustering is an incredibly powerful machine learning algorithm that expertly groups data points into clusters based on their similarities, in a way that each cluster has a parent cluster. Hierarchical clustering is a valuable technique that helps to group similar data points into clusters. Initially, each data point is considered as its cluster, and then the closest clusters are iteratively merged until all data points belong to a single cluster. There are two types of hierarchical clustering: agglomerative and divisive. The agglomerative method is the most commonly used type, where the algorithm merges the closest clusters until only one remains. On the other hand, the divisive method works by splitting the largest cluster into two smaller clusters until each data point belongs to its cluster [244].
- **BIRCH:** The BIRCH algorithm is an innovative hierarchical clustering algorithm that incorporates two fundamental concepts: Clustering Features (CF) and Cluster Feature Tree (CF Tree) to provide a more comprehensive cluster description. The CF Tree outlines the clustering of valuable information, and its minimal space requirement enables it to store metadata collections in memory, significantly enhancing the algorithm's speed and scalability. This makes it an ideal option for handling large datasets. It is beneficial for clustering both discrete and continuous attribute data [245]. According to [246], the first step in organizing the dataset objects involves creating a sub-clustering CF form. This form consists of a triple of information, denoted as  $CF = (N, LS, SS)$ , where  $N$  represents the number of data points,  $LS$  denotes the sum of the attribute values of  $X$ , and  $SS$  represents the sum of the squared values of  $X$ . The resulting CF is then clustered into  $k$ -groups using the conventional hierarchy clustering procedure. Should two CFs be merged, the theorem applies accordingly.

$$CF12 = (N1 + N2, \overline{LS}_1 + \overline{LS}_2, SS1 + SS2) \quad (47)$$

BIRCH is a meticulously designed algorithm that generates a concise summary of CF sub-clusters incrementally. The clusters are represented by a vector CF, which is the only value stored in memory. This CF value is sufficient to compute vital information related to subclusters, including their centroid, radius, and diameter. By

summarizing the information about subclusters instead of saving all points, BIRCH offers a highly efficient storage technique. The D2 distance formula is utilized to locate a cluster feature suitable for combination.

$$D2 = \frac{\sqrt{(N_1SS_1) + (N_2SS_2) + 2LS_1LS_2}}{N_1N_2} \tag{48}$$

By applying the following formula, we can determine the radius of a CF leaf.

$$R = \frac{\sqrt{SS - (LS)^2/n}}{n} \tag{49}$$

- Fuzzy C Means clustering (FCM):** This is an extremely powerful and flexible soft clustering algorithm that enables a data point to belong to multiple clusters with varying degrees of membership. This sets it apart from traditional clustering algorithms like k-means clustering, which only allows each data point to belong to a single cluster [247]. According to [248], the Fuzzy C-Means clustering algorithm computes the probability of image pixel membership to image clusters. In traditional Fuzzy C-Means clustering, the objective function to be minimized is:

$$J_m = \sum_{i=1}^D \sum_{j=1}^M u_{ij}^m \|x_i - c_j\|^2 \tag{50}$$

$$\text{With } \sum_{j=1}^M u_{ij} = 1 \tag{51}$$

Let's consider a scenario where we have an image X and want to group its pixels into clusters. In this case, we can represent the  $i^{th}$  pixel of the image X as  $x_i$ . Similarly, we can represent the  $j^{th}$  cluster center as  $c_j$ .  $u_{ij}$  is a representation of  $x_i$  degree of membership to the  $j^{th}$  center. M clusters and D number of picture pixels are present. Utilizing a natural number, m, we regulate the degree of fuzziness.

- Spatial Clustering:** Through a clustering analysis of observation points that exhibit comparable deformation sequences, the deformation area can be divided. To ensure alignment with spatial observation data clustering, essential similarity indicators, and a spatial similarity index have been established. These indicators include three primary spatial similarity measures: "weighted absolute distance," "weighted increment distance," and "weighted growth rate distance." These measures gauge the similarity between two distinct locations at different points in time.

The "weighted absolute distance" in real time between observation points k and l is denoted by  $d_{kl}^S(AD)$  and the full formula can be found here.

$$d_{kl}^S(AD) = \sum_{m=1}^M \sum_{t=1}^T WX_m [x_{mt}(k) - x_{mt}(l)]^2 \tag{52}$$

The value of the  $m^{th}$  deformation variable of the observation point k at the time section t ( $m = 1, 2, \dots, M, t = 1, 2, \dots, T$ ) is represented as  $x_{mt}(k)$ ,  $x_{mt}(k) = \delta_{mt}(k)$ ,  $x_{mt}(l) = \delta_{mt}(l)$ . In this case, the weight of the  $m^{th}$  deformation variable  $x_m$  is provided as  $WX_m$ . The distance between deformations at observation locations k and l at a given time instance is measured by the value of  $d_{kl}^S(AD)$ . The similarity between the deformation at the two observation places increases with decreasing  $d_{kl}^S(AD)$  value.

The "Weighted Increment Distance" in real time between observation locations  $k$  and  $l$  is denoted by  $d_{kl}^S(ID)$ , and the entire formula will be found here.

$$d_{kl}^S(ID) = \sum_{m=1}^M \sum_{t=1}^T WX_m [y_{mt}(k) - y_{mt}(l)]^2 \tag{53}$$

The equation in this case is  $y_{mt}(k) = x_{mt}(k) - x_{m,t-1}(k)$ ;  $y_{mt}(l) = x_{mt}(l) - x_{m,t-1}(l)$ . The deformation increment distance between observation points  $k$  and  $l$  at a particular time from the last measurement is indicated by the value of  $d_{kl}^S(ID)$ . The deformation increments at the two observation places are more likely to be identical if  $d_{kl}^S(ID)$  is lower.

$d_{kl}^S(GRD)$  represents the full-time "Weighted growth rate Distance" between observation sites  $k$  and  $l$ , and the full formula, if provided, is

$$d_{kl}^S(GRD) = \sum_{m=1}^M \sum_{t=1}^T WX_m [z_{mt}(k) - z_{mt}(l)]^2 \tag{54}$$

$$z_{mt}(k) = \frac{y_{mt}(k)}{x_{m,t-1}(k)} \tag{55}$$

$$z_{mt}(l) = \frac{y_{mt}(l)}{x_{m,t-1}(l)} \tag{56}$$

Here, the difference between the two observation points,  $k$ , and  $l$ , at one particular time and the last can be used to calculate the value of  $d_{kl}^S(GRD)$ , which indicates the relative deformation increments of the two points. A greater degree of remarkable similarity between the relative deformation increments of the two observation locations is indicated by a smaller value of  $d_{kl}^S(GRD)$  [249].

- **Spectral Clustering:** Spectral Clustering is a powerful technique for grouping data according to their similarities. It operates by assessing the interconnections between each data point and creating a graph with vertices representing the data and edges representing those relationships. Typically, the connections are determined by measuring the distance between two related records. Spectral clustering is a straightforward algorithm that can be efficiently solved using standard linear algebra software. Due to its superior performance compared to traditional clustering approaches like the k-means algorithm, Spectral Clustering has gained significant popularity in recent times [250].
- **Means Shift Clustering:** Means shift clustering is a clustering algorithm that groups data points based on their density in the feature space. Unlike parametric algorithms, it makes no assumptions about the data distribution. The algorithm selects a data point and defines its neighborhood using a kernel function that assigns a weight to each data point based on its distance from the current data point. Then, it shifts the selected data point towards the mean of the data points in its neighborhood, and the process is repeated for all data points until convergence is achieved. In simpler terms, mean shift clustering identifies clusters of data points based on how close they are to each other in space, without assuming any specific shape or size for the clusters [251]. The means shift clustering algorithm iteratively shifts each data point towards the mean of the data points in its neighborhood until all data points have converged to a local maximum of the density function. These local maxima represent the clusters in the data. Mean

shift clustering is particularly useful for clustering data with overlapping clusters or is not well-defined. It can also handle data with outliers, disrupting the performance of other clustering algorithms [252].

- **DBSCAN:** The Density-Based Clustering (DBSCAN) approach solves the problem of identifying clusters in data with varying densities. The algorithm allocates clusters in dense regions of the data space while separating regions with lower point density as noise. DBSCAN works by defining a neighborhood around each point in the data space and requiring that a minimum number of points fall within that neighborhood. Clusters are formed by connecting points that satisfy this criterion, while lone points that do not meet the minimum threshold are classified as noise. In simple terms, DBSCAN identifies clusters of data points by looking for areas with high point density and separating them from areas with low point density. It is a helpful algorithm for identifying clusters in complex data sets where traditional clustering algorithms may fail [253]. DBSCAN is an algorithm that can identify clusters in data sets with varying densities and classify outlier points as noise. It is beneficial for handling large spatial datasets with small related clusters in multiple dimensions, which can significantly reduce computation time. The algorithm evaluates the density of data points in a given space, grouping them based on their proximity. Points alone in low-density regions are classified as outliers, meaning they are not part of any cluster. It's worth noting that DBSCAN requires some adjustment for certain types of data sets to identify cluster shapes accurately. However, it remains a powerful clustering method that can effectively handle complex data sets [254].

### 7.1.3 Semi-supervised machine learning

Semi-supervised machine learning classification is an algorithm for categorizing data into different groups using a small amount of labeled data and a large amount of unlabeled data. This method differs from supervised machine learning classification, which relies solely on labeled data, and unsupervised machine learning classification, which relies only on unlabeled data. Semi-supervised machine learning classification is beneficial for problems with scarce or expensive labeled data. This is often the case in hyperspectral imaging, where the images can be large and complex, and it can be difficult and costly to label all of the pixels in the image [255].

- A. **Inductive SVMs:** According to [256], Inductive Support Vector Machines (SVMs) are well-suited classification algorithms for high-dimensional classification tasks. These algorithms aim to maximize the margin between the closest training samples for two classes by utilizing hyperplanes. The algorithm obtains the separating hyperplane by maximizing the separating margin between the two classes. This makes it an ideal choice for remote sensing classification problems.

Consider a set of training examples  $S = (x_i, y_i)$ , where  $i$  ranges from 1 to  $l$ , i.e.,  $i = 1, 2, \dots, l$ . Each input pattern  $x_i$  is associated with a label  $y_i$  that belongs to the set  $y_i \in \{\pm 1\}$ . The SVM classifier aims to minimize the error by using a nonlinear mapping  $\phi(\cdot)$ .



$$J(W, \epsilon) = \frac{1}{2} \|W\|^2 + C \sum_{i=1}^l \epsilon_i \tag{57}$$

Subject to:

$$y_i(\Phi(x_i) \cdot W + b) \geq 1 - \epsilon_i$$

Here,  $\epsilon_i = 1, 2, \dots, l$

Constrained Quadratic Programming (QP) is a proven and reliable method for confidently minimizing equations (58) and achieving significant reductions in both VC dimension and misclassification error.

The proposed solution yields a decision function that can be expressed in the subsequent format:

$$f(x) = \text{sgn} \left[ \sum_{i=1}^l y_i \alpha_i k(x, x_i) + b \right] \tag{58}$$

The function  $k(.,.)$  is defined as follows:

$$k(x, x_i) = \langle \Phi(x), \Phi(x_i) \rangle \tag{59}$$

Only a tiny fraction of the  $\alpha_i$  coefficients are non-zero, and the corresponding pairs of  $x_i$  entries are referred to as support vectors. These support vectors fully define the decision function. The term  $k(x, x_i)$  is the corresponding nonlinear kernel function.

For the experiment, the RBF kernel function in the form  $k(x, x_i) = \exp(-\gamma \|x_i - x_j\|^2)$  was used. This kernel function is defined by a weight  $c$ . The two-class SVM can be extended to multi-class classification by designing several one-against-all (OAA) two-class SVMs.

**B. Transductive SVM:** Semi-supervised learning algorithms, such as transductive support vector machines (SVMs), are an effective tool for classification tasks. These algorithms can assist in streamlining the classification process, making it more efficient and accurate. These algorithms can train a classification model by utilizing a small amount of labeled data and a large amount of unlabeled data. The transductive SVM method involves constructing a graph where each node represents a data point and the edges represent their similarity. Using this graph, the algorithm propagates the labels from the labeled data points to the unlabeled data points.

In [256], the transductive SVM is an iterative algorithm that gradually searches for an optimal separating hyperplane in the feature space. It does this through a transductive process incorporating unlabeled samples during training. In the semi-supervised framework, two datasets are defined: a labeled training dataset  $S$  and an unlabeled dataset  $V = [(x_j)], j = 1 + 1, \dots, n$ . The learning process of the TSVM can be formulated as an optimization problem as follows:

$$J(W, \epsilon, \epsilon^*) = \frac{1}{2} \|W\|^2 + C \sum_{i=1}^l \epsilon_i + C^* \sum_{i=1}^l \epsilon_j^* \tag{60}$$

Subject to:

$$y_i(\Phi(x_i) \cdot w + b) \geq 1 - \epsilon_i, \epsilon_i \geq 0; i = 1, 2, \dots, l$$

$$y_j(\emptyset(x_j).w + b) \geq 1 - \epsilon_j^*, \epsilon_j^* \geq 0; j = 1, 2, \dots, l$$

For the training and testing samples, the user-specified penalty levels are indicated by  $C$  and  $C^*$ , respectively. The number of transductive samples is denoted by  $d$ , and the slack variables are represented by  $\epsilon_i$  and  $\epsilon_j^*$ . The aforementioned optimization challenge must be resolved to train the Transductive Support Vector Machine (TSVM). Once the Lagrange multipliers  $\alpha_i$  and  $\alpha_j^*$  are set, the TSVM's decision function can be found.

$$f(x) = \text{sgn} \left[ \sum_{i=1}^l y_i \alpha_i k(x, x_i) + \sum_{j=1}^d y_j^* \alpha_j^* k(x, x_j^*) + b \right] \tag{61}$$

**C. Graph-based methods:** Graph-based methods are a powerful tool for classifying data. Constructing graphs with nodes representing labeled and unlabeled data samples and edges representing their similarities is an effective way to classify data samples. By propagating each sample's label information to its neighboring samples until a global stable state is reached, confident propagation of each data sample's label to its neighboring points becomes possible. This approach is highly effective for data classification tasks. In [257], According to the authors, the graph structure is represented as  $G = (V, E)$ , where  $V$  stands for the dataset's labeled and unlabeled data samples and  $E$  for the similarities between them.  $X = [x_1, x_2, \dots, x_M]$  represents the HSI dataset, where  $x_i$  is a member of  $FN$ . Let's have a look at this dataset. The feature vector is denoted by  $F$ , the total number of pixels in the HSI by  $M$ , and the total number of spectral bands, or feature dimension, by  $N$  in this case. Let  $U = \{1 + 1, \dots, 1 + u\}$  represent the unlabeled samples and  $L = \{1, \dots, l\}$  represent the labeled samples corresponding to labels  $y_1, \dots, y_l$ . We take two steps to build the graph. Using the  $k$ -nearest or  $e$ -nearest neighbor approach, we build the graph adjacency matrix in the first step. Using one of the following equations, we determine the graph weights in the second stage:

- The Gaussian similarity function, whose representation is as follows, is one of the formulas used to calculate graph weights.

$$g(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right) \tag{62}$$

Here, the  $\sigma$  factor controls the width of the neighbourhood

- The Gaussian similarity function, whose representation is as follows, is one of the formulas used to calculate graph weights.

$$g(x_i, x_j) = \|x_i - x_j\|^{-1} \tag{63}$$

Here,  $x_i$  and  $x_j$  are associated with weight  $w_{ij}$ . If samples are unconnected,  $w_{ij} = 0$ .

To enhance the ease of categorization, the weight matrix  $W$  is computed for all labeled and unlabeled data. The normalized graph Laplacian is precisely defined as:

$$L = I - D^{(-1/2)} W D^{(-1/2)} \tag{64}$$

Here,  $D$  is represented as diagonal matrices with degrees  $d_1, d_2, d_3, \dots, d_N$  and  $d_i = \sum_{j=1}^n w_{ij}$

The Laplacian is a powerful tool, possessing an essential property that drives innovation and progress, and it gives below,

$$F'LF = \frac{1}{2} \sum_{i,j=1}^n w_{ij} \left( \frac{f_i}{\sqrt{d_i}} - \frac{f_j}{\sqrt{d_j}} \right)^2 \quad (65)$$

Here, Vector F comprises various elements. The following objective function is intended for data classification and therefore should be minimized. Graph-based techniques are becoming increasingly popular among researchers for their sparse properties, robust mathematical basis, connection to kernel methods, and exceptional performance. In the following section, we will delve into some of the graph-based techniques utilized for HSI classification.

- D. Object-based classification:** Object-based classification (OBC) is a method of image classification that segments an image into objects and then classifies those objects based on their spectral, geometric, and spatial properties. Unlike traditional pixel-based classification methods, OBC does not classify each pixel in the image independently. OBC is useful for identifying and classifying objects with high precision in hyperspectral imaging due to the high spectral resolution of the images. C is also useful for classifying images with complex textures or mixed pixels, wherein traditional pixel-based classification methods can face challenges [258].
- E. Sub-Pixel-Based Classification:** Sub-Pixel-Based Classification (SPC) is a type of image classification that can identify and quantify materials in an image at a sub-pixel level. Unlike traditional pixel-based classification methods, SPC does not assign each pixel in the image to a single class. SPC is made possible by the high spectral resolution of hyperspectral photographs, which contain details about an object's spectral reflectance at hundreds or even thousands of tiny, contiguous wavelength bands. This allows researchers to identify and quantify materials in an image, even when mixed with other materials at the pixel level. Several SPC algorithms can be used [259].
- F. Super-Pixel-Based Classification:** Super-Pixel-Based Classification (SPBC) is an image classification technique that groups pixels into super-pixels before performing classification. By implementing this approach, we can achieve remarkable results with utmost accuracy and clarity, making a positive impact on our goals. Super-pixels are groups of pixels that are similar in color and texture. Using a graph-based algorithm to segment the image into regions based on pixel similarity is how image segmentation is typically achieved. Once the image has been segmented into super-pixels, a range of classification algorithms can be used to classify the super-pixels [260]. The combination of pixels with spatial proximity and spectral similarity in hyperspectral images is called super-pixel. Super-pixel classification is mainly used for segmentation. Super-pixel segmentation is utilized to extract spectral data from hyperspectral images, effectively reducing the number of units that must be classified and minimizing the impact of noise. It is a well-established fact that each segment can effectively be viewed as a super-pixel which serves as a crucial component of an object. Over-segmentation, on the other hand, is a widely recognized approach that enables the generation of super-pixels for representing local information and taking full advantage of the spatial correlation [261].

## 7.2 Neural networks

Neural network classification models are machine learning models that enable data classification. They are designed to mimic the structure and function of the human brain, consisting of interconnected nodes or neurons. Each neuron performs a simple mathematical operation, and the results are passed on to other neurons in the network. Neural network classification models are versatile and can be trained on different data types, such as images, text, and numbers. To train such a model, you provide labeled data where each point carries a known label. The neural network then learns to predict new data labels based on the patterns it has detected from the training data [262]. Neural network classification models are divided into two types:

- Traditional neural networks,
- Deep Learning

### 7.2.1 Traditional neural network

According to [263], a traditional neural network refers to a class of artificial neural networks designed to mimic the structure and function of the human brain. It consists of a series of layers of interconnected nodes, or neurons, where each neuron is connected to every neuron in the previous and the next layer. The input is fed to the first layer, and the output is obtained from the last layer. The neurons in each layer perform a simple mathematical operation on their input and pass their output to the next layer.

**Artificial neural network (ANN)** The Artificial Neural Network (ANN), takes inspiration from the structure and operation of biological neurons. It is a complex, multilayered system that can learn and extract numerous features. The network is composed of an input layer, multiple hidden layers, and an output layer to produce the final result. The computation process follows a specific format. The  $j^{\text{th}}$  neuron in the  $i^{\text{th}}$  layer is represented  $v_{ij}$ . The value has a certain form and is calculated using the neurons in the layer above.

$$v_{ij} = \Phi \left( v_{ij} + \sum_{k=1}^{n_{i-1}} w_{(i-j)k}^{ij} v_{(i-1)k} \right) \quad (66)$$

Here, the number of neurons in the  $(i - 1)^{\text{th}}$  layer is given by  $n_{i-1}$ , the connecting weights between the  $v_{ij}$  and  $v_{(i-1)k}$  neurons are depicted as  $w_{(i-j)k}^{ij}$ , and the bias index for the  $v_{ij}$  the neuron is given as  $b_{ij}$ . The pointwise activation function is given as  $\Phi(\cdot)$ , which is used to apply the non-linearity to the neural network [264].

- FNN:** The term FNN stands for Feedforward Neural Network, which is a type of artificial neural network that facilitates the flow of information in a unidirectional manner, starting from the input layer, traversing through the hidden layers, and eventually reaching the output layer. The input data is fed to the input layer, and each neuron in the input layer is connected to every neuron in the first hidden layer. The neurons in the hidden layers perform a simple mathematical operation on their input and pass their output to the next layer until the output layer produces the final output. FNNs are primarily used for supervised learning tasks such as classification and regression [264].
- MLP:** Unlock the power of neural networks with the Multi-Layer Perceptron, also known as MLP. Experience the wonder of multiple neuron layers working together in perfect harmony

to create a feedforward neural network. Each of these layers is connected to every neuron in the previous and next layers. MLPs are widely used in supervised learning tasks such as classification and regression. They are trained using backpropagation, which adjusts the weights of the connections between neurons to minimize the error between predicted and actual outputs. They effectively solve complex problems, especially non-linear relationships between input and output variables [265]. According to [266], the Multilayer Perceptron (MLP) is a type of feedforward artificial neural network where nodes from different layers are interconnected. It was first introduced by Frank Rosenblatt in his perceptron program. The perceptron is considered the basic unit of an artificial neural network and it defines the artificial neuron in the network. It is a supervised learning algorithm containing nodes' values, activation functions, inputs, and weights to calculate the output. The MLP neural network works only in the forward direction. All nodes are fully connected to the network. Each node passes its value to the next node only in the forward direction. The MLP neural network uses the backpropagation algorithm to improve the accuracy of the training model.

**Structure of MLP** This neural network comprises three main layers that work together to form an Artificial Neural Network.

**Input layer** This layer represents the output of the Neural Network. The number of nodes in the output layer depends on the problem type. For a single targeted variable, use one node. N classification problem, ANN uses N nodes in the output layer.

**Hidden layer** The hidden layer is responsible for all computations within the neural network. The edges of this layer are assigned weights, which are then multiplied by the node values. Additionally, the hidden layer utilizes an activation function. The model can have one or two hidden layers. It is essential to have several hidden layer nodes to achieve accuracy. Having too few nodes in the hidden layer can make the model inefficient in processing complex data. Conversely, having too many nodes can result in an overfitting problem.

**Output layer** The output layer of a Neural Network is responsible for providing the predicted output. The number of nodes required in this layer depends on the type of problem being solved. For a problem where only one variable is being predicted, one node is sufficient. However, for an N-classification problem, the output layer should have N nodes to facilitate the classification process (Fig 21).

## 7.2.2 Deep learning

Deep learning is the cutting-edge subset of machine learning that is specifically designed to train deep neural networks with multiple layers, making it a powerful tool for solving complex and challenging problems. It is an artificial intelligence technique that enables systems to learn and improve from experience without being explicitly programmed. Deep learning algorithms are designed to identify patterns and relationships in large, complex datasets, and they have proven to be highly effective in tasks such as image and speech recognition, natural language processing, and decision-making [267]. In deep learning, we have different types of classification models available: Autoencoder, Attention Models, Transformer models, etc., are available. We have divided the CNN models into subsections. The description of these models is given below

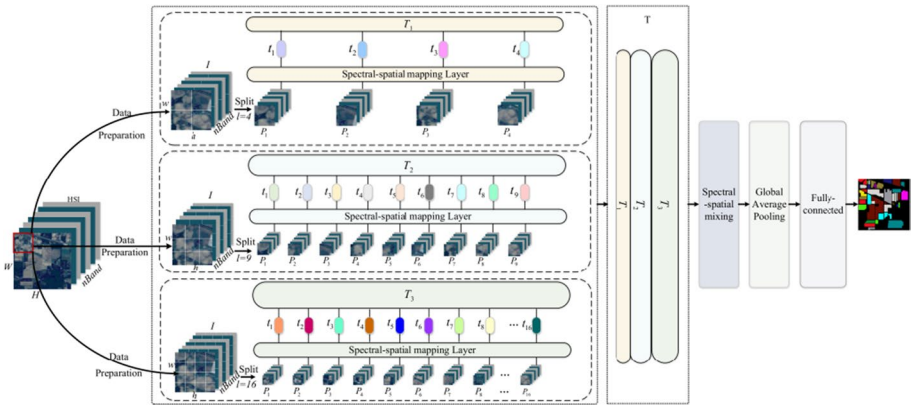


Fig. 21 The framework of the Multiscale-MLP for HSI classification [266]

A. **Autoencoders:** Autoencoder is a powerful tool in the field of artificial neural networks, particularly in its ability to learn efficient data representations. These models are unsupervised learning models, meaning they do not require labeled data to train. Autoencoders are also used as unsupervised dimensionality reduction techniques. It is used to learn a mapping from high-dimensional observations to low-dimensional representation space. The original observation can be reconstructed from the lower-dimensional representation [268], which is used for image pre-processing, feature extractions, and image classification. According to [29], it minimizes input and reconstructed output differences. Autoencoders have the encoder and decoder and the reconstructed output. These are the visible input layers called  $x$ , hidden layer  $h$ , reconstruction layer of  $x$  units, and activation layer  $f$ . An autoencoder is a feedforward technique to reconstruct an output from input. For an input vector  $x$ , the “encoder” maps the input to a hidden layer and produces  $y$ . After that, we can get the encoded value  $y$  by the parameter weight  $w_y$  and bias  $b_y$ .

$$y = f(w_y + b_y) \tag{67}$$

The "decoder" is responsible for mapping  $y$  to an output layer that is of the same size as the input layer. This output layer is commonly referred to as "reconstruction".

$$z = f(w_z + b_z) \tag{68}$$

The hidden layer to output weights and the input-to-hidden layer are represented by  $w_y$  and  $w_z$  in this instance. The activation function is displayed by  $f(\cdot)$ , and the bias of the hidden and output units are, respectively,  $b_y$  and  $b_z$ . The sigmoid, tanh, and rectified linear functions are just a few options for activation functions. In this work, we have used the sigmoid function as the activation function for both the encoder and decoder. It is defined in the below equation.

$$f(x) = \frac{1}{1 + e^x} \tag{69}$$

To achieve the objective of training, it is crucial to minimize the "error" between the input and output which is also commonly known as reconstruction. For this purpose, the

loss function is defined in a precise manner to ensure accurate and efficient performance during training.

$$J(\Theta) = \frac{1}{2M} \sum_{m=1}^M |z^m - x^m|^2 \tag{70}$$

The total number of training samples available is denoted by  $M$  in this case. It may be possible to successfully reduce the difference between the input and reconstructed output over the entire training set  $X = \{x_1, x_2, x_3, \dots, x_M\}$  to determine the values of  $\Theta = (w_y, w_z, b_y, b_z)$ .

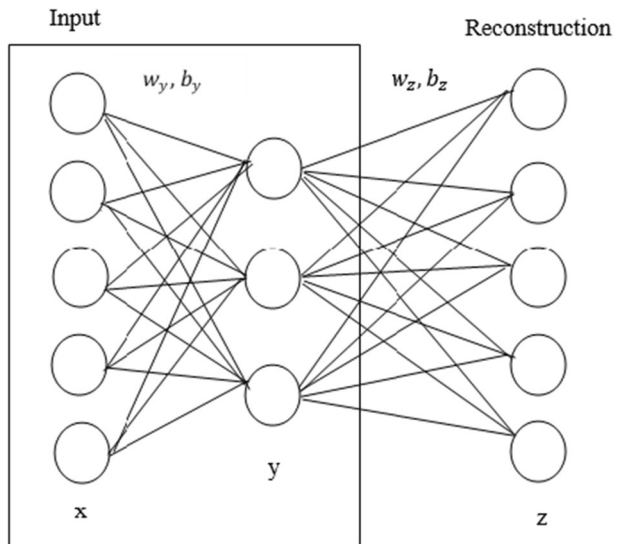
Through this kind of training focused on reconstruction, the authority and efficacy of AE are determined. It only uses the data from the hidden layer, represented as input features, during the reconstruction phase. This model must be able to perfectly recover the original input from  $y$  to demonstrate that it retains sufficient knowledge of the input (Fig. 22).

- B. **RNN:** According to [269], An RNN or Recurrent Neural Network is an artificial neural network that includes loops in connections, unlike a conventional feedforward neural network. These loops enable RNNs to handle sequential inputs using a recurrent hidden state that depends on the activation of the previous step. As a result, the network is capable of displaying dynamic temporal behavior. If we have given a sequence of data  $x = (x_1, x_2, \dots, x_T)$ , where  $x_i$  is the data at  $i^{th}$  timestep, an RNN updates its recurrent hidden states  $h_t$  by 0, if  $t=0$ ;  $h_t = \phi(h_{t-1}, x_t)$ . The equation includes a nonlinear function called  $\phi$ , which can be either a logistic sigmoid function or a hyperbolic tangent function. The RNN may also have a single output  $y_T$  for certain tasks like hyperspectral image classification. However, for some other tasks, the RNN may have multiple outputs

$$y = (y_1, y_2, \dots, y_T) \tag{71}$$

The recurrent hidden state update rule in equation (68) is commonly used in traditional RNN models.

Fig. 22 Autoencoder representation





$$h_t = \varnothing(Wx_t + Uh_{t-1}) \tag{72}$$

The coefficient matrix  $W$  is used in a conventional RNN model to calculate the input at the current time step. Conversely, recurrent hidden units at the preceding time step are activated using the coefficient matrix  $U$ . Based on an element's current state,  $h_t$ , an RNN can be used to construct a probability distribution for the subsequent element in a data sequence. An RNN's ability to capture a distribution over sequence data with changing length makes this feasible. The sequence probability,  $p(X_1, X_2, \dots, X_T)$ , can be broken down into

$$p(X_1, X_2, \dots, X_T) = p(X_1) \dots p(X_T|X_1, \dots, X_{T-1}) \tag{73}$$

Unlock the potential of each conditional probability distribution with the power of a recurrent network.

$$p(X_T|X_1, \dots, X_{T-1}) = \varnothing(h_t) \tag{74}$$

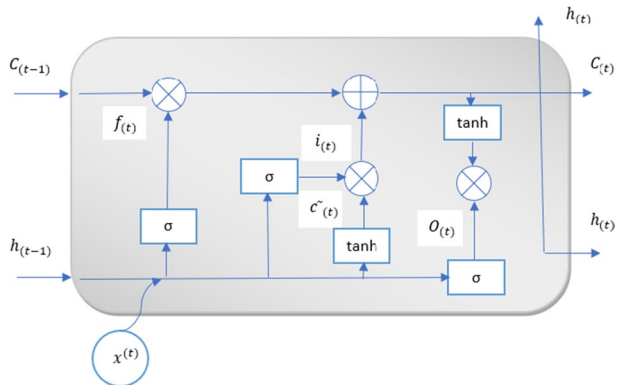
In this case,  $h_t$  is obtained from Eqs. (73) and (75). Since a hyperspectral pixel is treated as sequential data instead of a feature vector, we can use a recurrent network to model the spectral sequence. RNNs are an essential branch of the deep learning family. They have recently shown promising results in many machine learning and computer vision tasks. However, training RNNs to handle long-term sequential data can be challenging since the gradients tend to vanish. To address this issue, one common approach is to design a more sophisticated recurrent unit.

- C. **LSTM:** According to [270], to effectively solve the sequence learning problem, a Recurrent Neural Network (RNN) is the ideal choice as it incorporates recurrent edges that connect the neuron to itself across time, resulting in efficient and accurate learning outcomes. We have an input sequence  $\{x_1, x_2, \dots, x_T\}$ . And a sequence of hidden states  $\{h_1, h_2, \dots, h_T\}$ . At a given time  $t$ , the node with recurrent edge receives the input  $x_t$  and its previous output value  $h_{t-1}$ , then outputs the weighted sum of them, which can be formulated as below equation

$$h_t = \sigma(W_{hx}x_t + W_{hh}h_{t-1} + b) \tag{75}$$

The weight input node and the recurrent hidden node are represented here as  $W_{hx}$ , the bias is represented by  $b$ , and the non-linear activation function is represented by  $\sigma$ . The weight between the recurrent hidden node and itself from the previous time step is represented as  $W_{hh}$  (Fig. 23).

**Fig. 23** The architecture of LSTM



However, there's a difficulty with training RNN models. Depending on whether  $|W_{hh}| < 1$  or  $|W_{hh}| > 1$ , the contribution of the recurrent hidden node  $h_m$  at time  $m$  to itself  $h_n$  at time  $n$  may approach zero or infinity as  $n - m$  rises. Long-term dependencies pose a challenge for recurrent neural networks (RNNs) as back-propagating errors over many steps can lead to the gradient vanishing or exploding. To tackle this issue, a solution called long-term short memory (LSTM) was introduced. LSTM replaces the recurrent hidden node with a memory cell that stores and retrieves relevant information using dot product and matrix addition operations. This enables the network to better learn and remember long-term dependencies. The memory cell in LSTM has a node with a self-connected recurrent edge with a fixed weight, ensuring that the gradient can traverse numerous time steps without vanishing or exploding. LSTM comprises four crucial components: input gate, output gate, forget gate, and candidate cell value. By leveraging these components, we can compute the memory cell and output, and overcome the challenges of learning long-range dependencies in RNNs.

$$f_t = \sigma(W_{hf}.h_{t-1} + W_{xf}.x_t + b_f) \tag{76}$$

$$i_t = \sigma(W_{hi}.h_{t-1} + W_{xi}.x_t + b_i) \tag{77}$$

$$\hat{c}_t = \tanh(W_{hc}.h_{t-1} + W_{xc}.x_t + b_c) \tag{78}$$

$$C_t = f_t \circ C_{t-1} + i_t \tag{79}$$

$$O_t = \sigma(W_{ho}.h_{t-1} + W_{xo}.x_t + b_o) \tag{80}$$

$$h_t = O_t \circ \tanh(C_t) \tag{81}$$

Here,  $\sigma$  is the logistic sigmoid function, ‘.’ Is a matrix multiplication, ‘.’ is a dot product, and  $b_f, b_i, b_c$  and  $b_o$  are biased terms. The weight matrix subscripts have the apparent meanings. For instance,  $W_{hi}$  is the hidden input gate matrix,  $W_{xo}$  is the input–output gate matrix.

D. **Attention-based Models:** According to [271], to solve the bottleneck issue caused by a fixed-length encoding vector, which would restrict the decoder’s ability to access the input’s information. The dimensionality of their representation would be compelled to be the same as for shorter or simpler sequences, which is anticipated to become particularly problematic for long and complex sequences. The step-by-step computations of the alignment scores, the weights, and the context vector comprise Bahdanau et al.’s attention mechanism.

- **Alignment score:**  $h_i$  is the alignment model of the encoded states, and the previous decoder output is represented as  $S_{t-1}$ , to compute a score  $e_{t,i}$  is used. It displays how closely the input sequence’s elements match the position’s current output  $t$ . The alignment model is represented by a function  $a(.)$ , which can be implemented by using the feed-forward neural network

$$e_{t,i} = a(S_{t-1}, h_i) \tag{82}$$

- **Weights:**  $\alpha_{t,i}$  are the weights; these weights are computed by applying a SoftMax operation to the previously computed alignment scores:

$$\alpha_{t,i} = \text{softmax}(e_{t,i}) \quad (83)$$

- **Context vector:**  $C_t$  is represented as a unique context vector. It is used to feed into the decoder at each time step. A weighted sum of all computes it, T is depicted as encoder hidden states:

$$C_t = \sum_{i=1}^T \alpha_{t,i} h_i \quad (84)$$

The attention process can be reformulated into a universal form that can be applied to any sequence-to-sequence (abbreviated as seq2seq) action, even though the information may not necessarily be related sequentially.  $f_{att}(Q, K, V)$  is an example of an attention module; it operates on certain queries, keys, and values. It will produce a few weighted average vectors, Q, K, V, and  $\hat{V}$  in that order. After calculating the similarity score between Q and K, the weighted average vector over V is calculated. It is possible to formulate the weighted average vector using Eq. (84)

$$a_{i,j} = f_{sim}(q_i, k_j), \alpha_{i,j} = \frac{e^{a_{i,j}}}{\sum_j e^{a_{i,j}}} \quad (85)$$

$$\hat{V}_i = \sum_j \alpha_{i,j} v_j \quad (86)$$

In this case, the  $i^{th}$  key-value combination is  $V_j \in V$ , and the  $i^{th}$  query is  $q_i \in Q$ . A function called  $f_{sim}$  calculates each  $k_j$ , and  $q_i$  similarity score. Furthermore, the attendant vector for the query  $q_i$  is  $\hat{V}_i$ . The attention module creates a weighted average for each inquiry if Q and K/V are related to each other. The attention module nevertheless generates a weighted average vector even in the absence of permanent vectors, and this vector could include adaptive or superfluous information.

- Self-attention:** Self-attention models enable a neural network to concentrate on different parts of its output. This technique is beneficial for tasks like text summarization and machine translation. Self-attention models are frequently used in transformer-based architectures such as BERT and GPT-3 [272].
- Encoder-decoder attention:** Neural networks can focus on specific input data while producing output using encoder-decoder attention models. This method is valuable for tasks like image captioning and question answering. Encoder-decoder attention models are often used in sequence-to-sequence learning tasks such as text summarization and machine translation [273].
- Global attention:** Global attention models enable a neural network to consider all parts of the input data. This technique is beneficial for tasks like sentiment analysis and question answering. It is not uncommon for global attention models to be employed in both RNNs and CNNs [274].
- Local attention:** Local attention models enable a neural network to focus on a subset of the input data. This method is helpful for tasks like image classification and object detection. Local attention models are often used in CNNs [275].

- e. **Hierarchical attention:** Hierarchical attention models allow neural networks to focus on varying levels of abstraction within input data. This technique is beneficial for tasks like document classification and machine translation. Hierarchical attention models are often used in RNNs and CNNs. Apart from these general types of attention models, many specialized attention models are also developed for specific tasks. For instance, attention models designed for speech recognition, music generation, and video processing [276].

E. **Transformer Models:** According to [271], Transformers is a type of neural network architecture introduced in 2017 in the paper "Attention Is All You Need." They have become the go-to model for various Natural Language Processing (NLP) tasks, such as machine translation, text summarization, and question-answering. Transformers are based on the self-attention mechanism, which enables the network to learn long-range correlations in the input sequence. This makes them particularly well-suited for NLP tasks where the meaning of a word or sentence can often depend on words far away from the sequence (Fig. 24).

Transformers consist of an encoder and a decoder which respectively generate a hidden representation and output sequence (Fig. 25). The decoder confidently employs the hidden state representation generated by the encoder to produce the output sequence. In a Transformer network, both the encoder and decoder contain self-attention layers that facilitate the learning of long-range dependencies by attending to various parts of the input sequence. The Transformer architecture has exhibited impressive performance in a range of NLP tasks, including but not limited to machine translation, text summarization, and question-answering. They are also being used for other tasks, including image classification and speech recognition [277]. Also, in transformers, we have so many types like Vision Transformer (ViT), Swin Transformers, Bidirectional Encoder Representations from Transformers (BERT), GPT, etc.

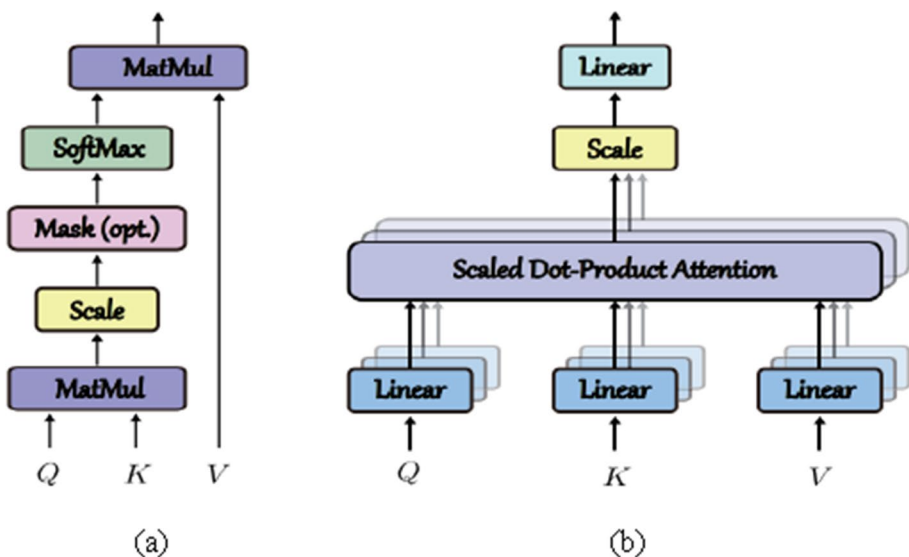
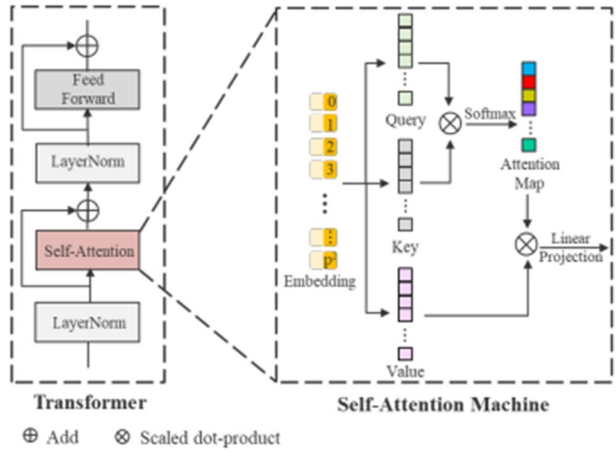


Fig. 24 Transformers serve as an example of the attention mechanism. a Self-attention module; b multi-head attention [277]

**Fig. 25** Transformer structure [278]



**Basics of transformers** According to [278], While RNNs have sequence dependence as a characteristic, Transformers adopt a self-attention mechanism that allows the network to efficiently capture global information, including long-term dependencies, for units at any position. This results in the complete abandonment of the aforementioned sequence dependence characteristic of RNNs. The use of transformers greatly improves the development of models for processing time series data. One should be aware that transformers can be applied to more than just NLP tasks, as they also have applications in image processing and computer vision. The vision transformer (ViT) has been proven or come close to achieving state-of-the-art results in a range of vision-related tasks, presenting exciting prospects for innovation and creativity in these domains. The success of transformers largely relies on multi-head attention, which involves stacking and integrating multiple self-attention (SA) layers. The self-attention mechanism is highly skilled at identifying and capturing the internal correlations within data or features. This means that it can lessen the reliance on external information. Figure 2(a) depicts the self-attention module in transformers. To perform self-attention, we can confidently follow these six easy steps:

- Step 1: Enter the  $m$ -long sequence data  $x$ , where  $x_i, i = 1, \dots, m$  is a vector or a scalar.
- Step 2: Using a shared matrix  $W$ , obtain the feature embedding, represented as  $a_i$ , for each scalar or vector  $x_i$ .
- Step 3: Get three vectors: Query ( $Q = [q_1, q_2, \dots, q_m]$ ), Key ( $K = [k_1, k_2, \dots, k_m]$ ), and Value ( $V = [V_1, V_2, \dots, V_m]$ ) by multiplying each embedding by three distinct transformation matrices,  $W_q, W_k$ , and  $W_v$ , respectively.
- Step 4: Calculate the attention score ( $s$ ) as an inner product, such as  $q_i \cdot k_j$ , between each  $Q$  vector and each  $K$  vector. To stabilize the gradients, the scaled score  $S_{i,j} = q_i \cdot k_j / \sqrt{d}$ , where  $d$  is the dimension of  $q_i$  or  $k_j$ , is calculated using normalization.
- Step 5: The Softmax activation function is applied on  $s$ .  $\hat{S}_{1,i} = e^{s_{1,i}} / \sum_j e^{s_{1,j}}$  is one example of this at position 1.
- Step 6: Develop attention representations  $z = [z_1, z_2, \dots, z_m]$ , where  $z_1 = \sum_j \hat{S}_{1,i} v_j$ . To sum up, the SA layer can be expressed as follows in its entirety:

$$z = \text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d}}\right)V \quad (87)$$

To create a multi-head attention model, we can combine multiple SA (self-attention) layers by using Eq. (88), as illustrated in Fig. 24(b). Firstly, we obtain several attention representations, which we denote as  $z^h, h = 1, 2, \dots, 8$ . These representations are obtained by applying SA multiple times (in this case, eight times). After generating the attention representations, we combine them to form a larger feature matrix. To ensure that the feature matrix has the same dimension as the input data, we use a linear transformation matrix. However, it's important to note that the Self-Attention layer doesn't incorporate any positional information, which means it fails to take into account the sequence information. To overcome this limitation, we encode the position information into the feature embedding. This embedding is formulated as  $a_i + e_i$ , where  $e_i$  is a unique positional vector that is manually given and represents the position of the embedding in the sequence.

"Vision Transformer" (ViT) is used to classify hyperspectral images. ViT is mainly applied in a straight line to Computer Vision (CV) with the most minor conceivable adjustments. Transformers can be used to extract the comprehensive evidence constructed on their unique structure to acquire long-range information [279]. The transformer model comprises undistinguishable layers, and each layer is composed of two sub-layers, named a fully connected feed-forward network and a multi-headed self-attention mechanism. Attention models represent global dependence inside a series of input transformers. The problem of vanishing gradient is a common problem for deep learning, including transformers. This impedes the training procedure's convergence [280].

**Convolutional neural network models** Hyperspectral image classification can be done using various types of CNNs. The most commonly used types are 2D CNNs, 3D CNNs, and spectral-spatial CNNs. 2D CNNs are the primary type of CNN used to classify images with two dimensions: height and width. However, 2D CNNs can also classify hyperspectral images by treating each spectral band as a separate channel [281]. 3D CNNs can classify data using height, width, and depth. Hyperspectral image classification can be improved with the use of spectral-spatial CNNs. These convolutional neural networks (CNNs) are expertly crafted to effectively leverage both spatial and spectral information present in images with precision. By combining the strengths of 2D and 3D CNNs, they can learn spatial and spectral features more effectively. Specialized CNN architectures have also been developed for hyperspectral image classification, including those for land cover classification, crop classification, and mineral mapping [282].

A. **Convolutional Neural Network (CNN):** According to [310], Stacks of convolutional, pooling, and nonlinear layers are used by a convolutional neural network to extract features. Using convolutional kernels, the convolutional layers calculate the convolution of input feature maps. Equation (89) gives the activity of the  $i^{\text{th}}$  feature map in the  $l^{\text{th}}$  layer.

$$y_i^l = \sum_j f(w_{ij}^l * y_j^{l-1} + b_i^l) \quad (88)$$

The feature map in layer  $l-1$  in this case is  $y_j^{l-1}$  and it is related to the feature map  $y_i^l$ .  $w_{i,j}^l$  is the convolutional kernel for  $y_j^{l-1}$ . The value of bias is  $b_{i,l}$ . The activation functions are rectified linear unit (ReLU) and sigmoid, and the non-linear activation function is denoted as  $f(\cdot)$ . The convolutional operator is indicated by a  $*$ . It is typical to incorporate a pooling layer after the convolutional layer. Max pooling is the most frequently used pooling function. It works by computing the maximum value in a particular window of the feature map. This process aids in making the feature map more robust against data distortion and enhances its invariance. Additionally, the pooling layer can decrease the feature map's size, reducing the computational burden.

By combining multiple convolutional and pooling layers, a sophisticated deep neural network is formed. This layered network can learn hierarchical features, with lower layers identifying fundamental features like edges and textures, and higher layers acquiring more intricate and abstract features with semantic significance. These features are highly advantageous in a wide range of applications, including classification tasks (Fig. 26).

- a. **2D-CNNs:** According to [283], The 2D-CNN technique applies a convolution process to the input data using 2D kernels. This involves computing the dot product between the input data and the kernel, which is then summed up. To ensure the entire spatial dimension of the input data is covered, the kernel moves over it. The features obtained from this process are subjected to an activation function to introduce non-linearity in the model. This technique is popularly used in image processing and computer vision applications, as it enables the extraction of crucial features from the input data. The activation value at spatial location  $(x, y)$  in the  $j^{th}$  feature map of the  $i^{th}$  layer in 2D convolution is computed using the following formula and is referred to as  $v_{i,j}^{x,y}$  (Fig. 27).

$$v_{i,j}^{x,y} = \phi \left( b_{i,l} + \sum_{\tau=1}^{d_{l-1}} \sum_{\rho=-\gamma}^{\gamma} \sum_{\sigma=-\delta}^{\delta} w_{i,j,\tau}^{\sigma,\rho} \times v_{i-1,\tau}^{x+\sigma,y+\rho} \right) \tag{89}$$

The activation function  $\phi$  is used in the preceding equation to determine the activation value at spatial point  $(x, y)$  in the  $j^{th}$  feature map of the  $i^{th}$  layer.  $b_{i,j}$  is the bias parameter for the  $j^{th}$  feature map of the  $i^{th}$  layer. The number of feature maps in the  $(I - 1)^{th}$  layer and the depth of the kernel  $w_{i,j}$  for the  $i^{th}$  layer's  $j^{th}$  feature map is represented by the values of  $d_{l-1}$ . In this case, the kernel's width is represented by  $2\gamma + 1$ , and its height by  $2\delta + 1$ . Last but not least, the value of the weight parameter for the  $i^{th}$  layer's  $j^{th}$  feature map is represented by  $w_{i,j}$ .

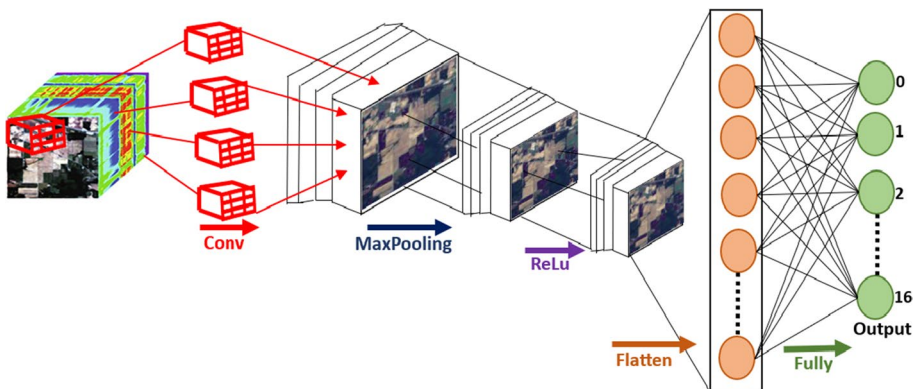
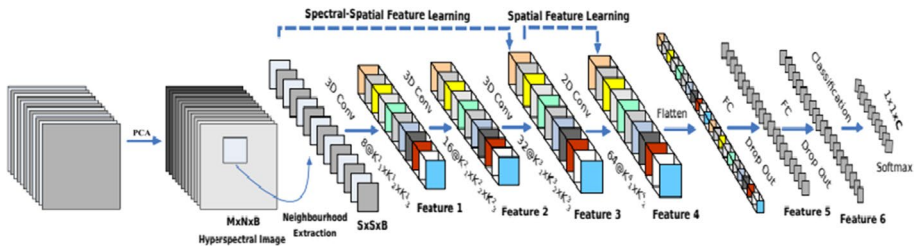


Fig. 26 CNN for hyperspectral image classification





**Fig. 27** HybridSpectralNet (HybridSN) Model, which integrates 3D and 2D convolutions for hyperspectral image (HSI) classification [283]

b. **3D-CNNs:** In 3D convolution, to effectively capture the spectral information in the 3D data, the proposed HSI data model utilizes a 3D kernel that is convolved with the data. The feature maps of the convolution layer are then created by applying this 3D kernel over multiple adjacent bands in the input layer. This approach has proven to be highly effective in capturing the spectral information in the data with confidence. The activation value at spatial location  $(x, y, z)$  in the  $j^{th}$  feature map of the  $i^{th}$  layer in 3D convolution is produced by applying the following equation to  $v_{i,j}^{x,y,z}$  (Fig 27)

$$v_{i,j}^{x,y,z} = \varphi \left( b_{i,l} + \sum_{\tau=1}^{d_i-1} \sum_{\lambda=-\eta}^{\eta} \sum_{\rho=-\gamma}^{\gamma} \sum_{\sigma=-\delta}^{\delta} W_{i,j,\tau}^{\sigma,\rho,\lambda} \times v_{i-1,\tau}^{x+\sigma,y+\rho,z+\lambda} \right) \quad (90)$$

The depth of the kernel along a spectral dimension in the provided equation is represented by  $2\eta + 1$ , and all other parameters stay the same as they did in the previous equation [284].

## 8 Hyperspectral image prediction

Hyperspectral image prediction refers to the process of predicting the properties of objects within a hyperspectral image. Hyperspectral images offer an unparalleled level of precision in identifying and classifying objects, thanks to their high spectral resolution and ability to capture hundreds or even thousands of narrow, contiguous wavelength bands [285]. The LU/LC changes are challenging the universal atmosphere variations significantly. The global atmosphere suffers from substantial fluctuations due to human activity at before intensities, unheard-of rates, and geographic scales. Human-induced land use transformation is a significant contributor and component of environmental diversity on a global scale. It is essential to possess knowledge of land use/cover and possible applications for choosing, planning, and implementing land use plans. Land use will impact the land cover, and the latter will affect the former [286]. SOC (Soil Organic Carbon) is a critical gauge of soil’s biological, chemical, and physical features in agricultural settings. It also makes up a significant portion of the world’s carbon cycle. Crop yields, a reduction in the ability of the soil to retain moisture, and an excess of nutrients can all arise from soil erosion. Geographically, dissimilarities in vegetation categories, soil features, and soil erosion rates fetched by deviations with slope inclination, the deepness of surviving crops, microclimate, slope processes, and soil physical characteristics are highly variable [287].

It is widely recognized that there are various approaches to predicting future land use/land cover (LU/LC) changes, which take into account factors such as the percentage and rate of change observed over a given period. The altered area is confidently examined to determine the differences between specific periods. With confidence, one may utilize both dependent and independent variables to examine the potential for LU/LC change. These variables include aspects such as distance and slope and can provide valuable insight into the environmental factors that may contribute to such changes [288]. Various techniques, including time neural networks, regression models, and series models, are employed to forecast changes in Land Use/Land Cover (LU/LC). These models consider factors such as water bodies and forest edges to make predictions about future changes. Figure 28 illustrates the range of methods utilized for predicting LU/LC changes. Additionally, there are numerous applications for predicting hyperspectral imaging, such as forecasting vegetation, soya bean growth, agriculture, and soil biochar levels.

LU/LC change prediction predicts how much land is used to build an area and how much land is occupied by water bodies. This will be used for urban planning and the developed area [289]. Vegetation prediction is used to predict the growth and disease of plants. If the plant's disease is predicted before, it will be suggested to give the fertilizers according to its growth. Soya bean prediction is used to indicate the soya bean crop, which will be used to know the plant's health. Soil prediction determines the soil's physical, chemical, and biological properties. If the soil properties are predicted before, an agriculture plan is to be done. We are concentrating on predicting the change in LU/LC from the applications mentioned above. Regression, neural network, and Time series models are prediction models for investigating future LU/LC changes [290].

## 8.1 Traditional machine learning-based prediction models

Traditional machine learning models for hyperspectral image prediction are statistical models trained on labeled hyperspectral data. In the field of hyperspectral data analysis, the training process confidently establishes the correlation between the spectral features present in the data and the respective target variable. This enables the model to predict the target variable for new hyperspectral data. Decision trees, random forests, Support Vector Machines (SVMs), and k-nearest Neighbors (k-NN) models are some of the commonly used traditional machine learning models for hyperspectral image prediction tasks [34]. Decision trees employ a tree-like structure to analyze data and make decisions

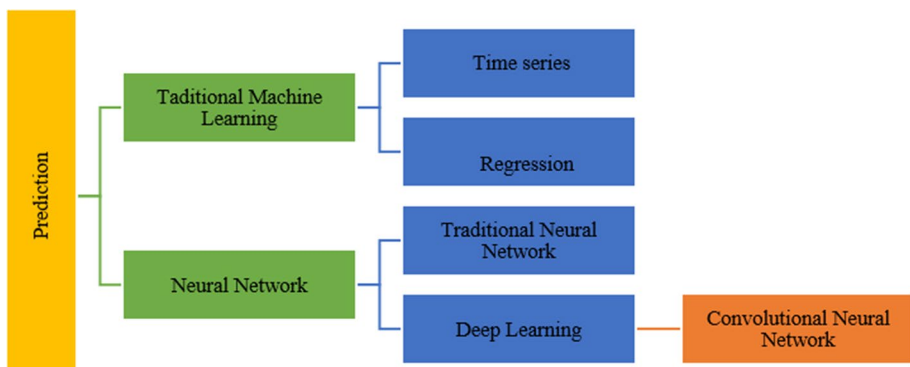


Fig. 28 prediction models for hyperspectral images

based on its spectral features. To classify and predict, two popular machine learning models are often used: random forests and SVMs. Random forests improve accuracy by ensembling multiple decision trees, while SVMs use a hyperplane to classify the data into distinct classes. K-NN models are a type of model that uses the spectral features of the nearest neighbors to make a prediction. These traditional machine learning models have strengths and can be used effectively for a range of hyperspectral image prediction tasks. By leveraging the power of these models, researchers and practitioners can gain insights into hyperspectral data and make accurate predictions about the target variable [291].

### 8.1.1 Time series models

Time series prediction involves forecasting future events based on past data with timestamps. It comprises developing models through historical research, using them to make judgments and direct future strategic decision-making. The following are the time series for hyperspectral LU/LC change prediction.

- a. **Markov Chain (MC):** Based on the transition probabilities, it is utilized to analyze the time-based changes in the landscapes among the LU/LC classes. Markov Chain (MC) investigation is a stochastic modeling technique that exploded widely employed to investigate the fluctuations of land use change at several balances. It operates based on physics assumptions, which recommend that if an organization's ailment at a previous time is recognized, the likelihood of it being in that state later can be calculated. It is my pleasure to suggest that, based on the analysis of the Markov Chain model, changes in Land Use and Land Cover (LU/LC) at a large scale can be predicted with a high degree of accuracy [292].
- b. **Cellular Automata (CA):** it is used to know the LU/LC changes to model the spatially evolving environments in the remote sensing environment. It is a dynamic bottom-up model. It is used to know the spatial dimension and model direction [128].
- c. **ARIMA (Autoregressive Integrated Moving Average Model):** Our powerful prediction tool utilizes time series data and advanced statistical analysis to confidently forecast future trends. By leveraging linear regressions, we can accurately predict future outcomes based on past data with a high degree of confidence. ARIMA is a prominent and commonly used statistical approach for predicting time series. It can capture several conventional temporal structures in time series data [293].

### 8.1.2 Regression models

Regression is a method for determining how independent traits or variables relate to a dependent feature or result. It is a machine learning predictive modeling technique using an algorithm to forecast continuous outcomes. One or more independent (predictor) variables and one or more dependent (criterion) variables are related in regression analysis. The criteria of the projected value is obtained from a linear combination of the predictors [294].

- a. **Linear Regression:** The link between factors impacting forest cover loss was determined using linear regression. GIS was used to create independent variables such as digital elevation data, distance from residential areas, distance from the road, and abruptness. A linear regression association was established between forest cover loss as a dependent variable and the stated factors [295].

- b. **Logistic Regression:** According to [295], Logistic Regression investigation ties the coincidental landslide (ranging from 0 to 1) to "u" logit (u0 indicates a higher likelihood of non-occurrence and 0u suggests a higher likelihood of occurrence). The logit "u" is considered to be a linear grouping of independent variables in logistic regression analysis, and the equations are as follows:

$$P_r = e^u / (1 + e^u) \quad (91)$$

Here, P is the model output, representing landslide occurrences probability, and u is the independent variable, a linear sum of factors (ex., land cover, slope, etc.).

## 8.2 Traditional neural network

Traditional neural network models are often used for hyperspectral image prediction tasks. By leveraging an interconnected network of nodes, our models confidently learn the intricate relationships between hyperspectral image features and the target variable for accurate predictions [296]. Several types of traditional neural network prediction models are commonly used for hyperspectral image prediction, including Multilayer Perceptron (MLPs), Radial Basis Function Networks (RBFNs), and Convolutional Neural Networks (CNNs). MLPs consist of a series of fully connected layers and are a simple yet effective type of neural network. RBFNs use radial basis functions as activation functions and are known for their ability to learn non-linear relationships in data. CNNs are known for their effectiveness in processing image data. Convolutional layers are a widely adopted approach for extracting features from data and predicting target variables. This technique involves the application of convolutional filters to the input data, which enables the detection of patterns and features at different scales. These traditional neural network models have unique strengths and can achieve state-of-the-art results for various hyperspectral image prediction tasks [112].

### 8.2.1 ANN-based models for prediction

Neural networks can also be used for prediction. Neural networks fix extrapolative analytics networks effectively because of their hidden layers. In linear regression models, only input and output nodes produce predictions. The neural network utilizes a hidden layer to improve prediction accuracy.

- a. **Back Propagation Neural Network (BPNN):** This model can incorporate the need to adjust activation functions. BPNN has double hidden layers, which outperformed the others regarding steadiness and simplification. The predicted accuracy was enhanced, although adding more hidden layers resulted in overfitting. The multi-layer BPNN model improved predicted accuracy and model stability by using a sophisticated weight calculation across hidden layers [297]. A boundless dimension is demonstrated to fit non-linearity among hyperspectral variables and soil nutrients. Because of how it manages failures, BPNN is known as a Multilayer Perceptron (MLP) network. Through propagating the output error back into the network, back-propagation overcomes the problem of "assignment of mistake in prediction to whatever input group." This method is continued until the input layer is reached with the lowest possible error model. A BPNN model typically involves input, output, and hidden layers [297].

- b. **Recurrent Neural Network (RNN):** Because they include sequences in the architectural design of network units, RNNs are more flexible for time series investigations. The vanishing gradient and expanding gradient difficulties make capturing lengthy temporal data with a typical RNN model problematic [298].
- c. **Long Short-Term Memory Neural Network (LSTM):** LSTM with particular hidden units was proposed for learning time series over extended periods. The capacity of LSTM to recall information in lengthy time series makes it worthwhile in various disciplines, including voice recognition, video analysis, and biology [298].

### 8.3 Deep learning models for prediction

Deep learning prediction models are powerful machine learning model that uses a deep neural network architecture to learn the relationships between the spectral features in a hyperspectral image and the target variable. These models comprise multiple layers of interconnected nodes, and each layer learns to extract more complex features from the data. Deep learning prediction models offer several advantages over traditional machine learning and neural network prediction models for hyperspectral image prediction. One advantage is that they can learn complex patterns in hyperspectral data without needing engineered features [122]. Additionally, they are less sensitive to the choice of hyperparameters and less likely to overfit the training data. Finally, deep learning prediction models can be trained on large datasets of hyperspectral images to achieve state-of-the-art results. Due to these advantages, deep learning prediction models are becoming increasingly popular for hyperspectral image prediction tasks. It has been observed that significant progress has been made in achieving commendable outcomes across a range of tasks, including classification, regression, and anomaly detection [299]. CNN-based prediction models are a deep learning model well-suited for hyperspectral image prediction tasks. These models can learn complex spatial and spectral features from hyperspectral images, allowing them to achieve state-of-the-art results on various hyperspectral image prediction tasks [300]. Several types of CNN-based prediction models are commonly used for hyperspectral image prediction, including 3D CNNs, spectral-spatial CNNs, Residual CNNs (ResNets), and Densely Connected Networks (DenseNets). 3D CNNs are a type of CNN architecture designed to process 3D data. These models can process hyperspectral images by treating each band as a separate channel. The Spectral-spatial CNN architecture is highly specialized for predicting hyperspectral images. It enables the CNN to learn both spatial and spectral features simultaneously from the hyperspectral data, which is a significant advantage. These models incorporate various techniques to combine spectral and spatial information from the hyperspectral data, leading to a substantial improvement in the model's performance [284].

## 9 Dataset description






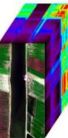



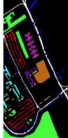

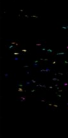
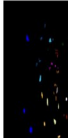


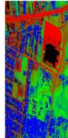
Hyperspectral image classification datasets are crucial for identifying and classifying objects accurately. These datasets consist of labeled hyperspectral images that provide detailed spectral information about objects at hundreds or even thousands of narrow, contiguous wavelength bands. With this high spectral resolution, researchers can differentiate between objects with similar characteristics and classify them with precision. To facilitate research and development, by using the following link we can download various types of benchmark datasets [https://www.ehu.eus/ccwintco/index.php/Hyperspectral\\_Remote\\_Sensing\\_Scenes](https://www.ehu.eus/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes).

1. **Indian Pines:** On June 12, 1992, in the region of North-western Indiana, advanced AVIRIS sensors were deployed to collect valuable data. The Airborne Visible/Infrared Imaging Spectrometer sensors helped to capture detailed information about the area which would have been difficult to obtain through traditional methods. This data set is divided into 16 classes. It consists of a 145\*145-pixel size; the spatial resolution of this dataset is 20 m. IP dataset owns 224 bands; after removing the water absorption bands, 200 bands are available. The wavelength of these images is 0.4~2.5. Table 9 Indian Pines Images are available before classification and ground truth images [301].
2. **Pavia University:** It was attained using the ROSIS (Reflective Optic System Imaging Spectrometer) sensors from a site in Northern Italy in 2002, 8 July. The Pavia University dataset has two types: one is Pavia University, and the other one is Pavia Centre [88]. Table 9 shows Pavia University Images are available before classification and ground truth images.
3. **Salinas:** According to [302], Salinas's dataset is collected from Salinas Valley, California. It was collected using AVIRIS sensors on 8 Oct 1998. This dataset also has two types: the Salinas-A and the Salinas scene. Table 9 shows Salinas Images before classification and ground truth images.
4. **Botswana:** According to [303], this dataset is collected using NASA EO-1 sensors in Okavango Delta, Botswana. This has 145 bands with a pixel resolution of 1476\*256. Table 9 shows Botswana Images are available before classification and ground truth images.
5. **Houston:** The Houston dataset is collected with CASI-1500(Compact Airborne Spectrographic Imager) sensors in Houston, USA. And this has nine classes and 144 bands [303]. In Table 9, Houston Images are available before classification and ground truth images.
6. **Kennedy Space Centre:** This dataset is collected from the Kennedy Space Center in Florida using NASA AVIRIS sensors. This has 176 bands and 13 classes.512\*614 pixels in the wavelength 400-2500 nm electromagnetic spectrum [302]. In Table 9, Kennedy Space Centre Images are available before classification and ground truth images.
7. **WHU-Hi-LongKu: The WHU-Hi-LongKu** dataset is collected from Hubei Province, China, using Headwall Nano-spectrometer. This has nine classes and 270 bands. Table 9 shows WHU-Hi-LongKu images before classification and ground truth images [88].
8. **HYDICE:** The dataset for the Hyperspectral Digital Imagery Collection Experiment (HYDICE) was taken from the mall in Washington, DC. This has 210 bands with 2.8 m spatial resolution. The spectral region is 0.4–2.4  $\mu\text{m}$ . These images have 304\*301 pixels and classes [304]. Table 9 shows HYDICE dataset Images before classification and ground truth images.
9. **Gaofen-5 (GF-5) Advanced Hyperspectral Imager (AHSI):** This land cover was launched on May 9, 2018 in China. It consists of mixed landscapes, including urban-rural outlying and mining extents. The spectral range is 400-2500 nm, and the spatial range is 30 m. It contains 330 bands, including six land cover classes [305] (Table 9).

## 10 Quality metrics

Accuracy is a metric used to determine which model is better for identifying relationships and patterns between variables in a dataset using input or training data. It is used for evaluating the classification models. The grade of concordance between the classification outcomes and the real-time appearances is evaluated using several methods. The dataset has to be split into test sets and train sets before the classification accuracy calculation can be

**Table 9** Description of benchmark datasets

Data	Indian Pines	Pavia University	Salinas	Botswana	Kennedy Space Centre (KSC)	WHU-Hi-LongKou	Houston	Hydice
Collection Location	Indiana, USA	Northern Italy	Salinas Valley, California	Okavango Delta, Botswana	Kennedy Space Centre, Florida	Hubei Province, China	Houston, USA	Washington, DC
Acquisition Equipment	AVIRIS	ROSIS	AVIRIS	NASA EO-1	NASA AVIRIS	Headwall Nano-Hyperspec	CASI-1500	
Number of Categories	16	9	16	14	13	9	9	6
Number of Bands	224	103	224	145	176	270	144	210
Number of bands after denoising	220	103	220	NA	NA	NA	144	NA
Data Size (Pixel)	145*145	610*601	86*63	NA	512*614	550*400	349*1095	304*301
Spatial Resolution (m)	20	1.3	3.7	30	NA	0.463	2.6	2.8
Spectral Coverage (µm)	0.4~2.5	NA	0.4~2.5	NA	0.4~2.5	NA	0.38~1.09	0.4~2.4
Sample size	10249	42776	54129	NA	NA	204542	15029	NA
Dataset								
Ground Truth								

done. Upon completion of training with datasets and subsequent testing, the resulting output is appropriately categorized. It is imperative to note that the accuracy and effectiveness of the model are heavily reliant on the quality and integrity of the datasets used for training. Therefore, it is crucial to use high-quality datasets to achieve accurate results [306]. It won't be possible to know the accuracy of the trained model until after it has been evaluated. To assess the proposed model's performance with the other existing models, the commonly used metrics are Confusion Matrix (CM), Average Accuracy (AA), Kappa Coefficient (KC), Overall Accuracy (OA), NDVI, NDWI, and rate and percentage of change, the following are the formulas for CM, OA, AA, KC, NDVI, NDWI, rate, and percentage of change.

1. **Confusion Matrix:** According to [228], The confusion matrix, also known as an error matrix, is mostly utilized for comparing the original ground cover's categorization conclusion. Given that the confusion matrix's order is  $c \times c$ ,



$$X = \begin{bmatrix} x_{11} & x_{12} & \dots & \dots & \dots & x_{1c} \\ x_{21} & x_{22} & \dots & \dots & \dots & x_{2c} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ x_{c1} & x_{c2} & \dots & \dots & \dots & x_{cc} \end{bmatrix}$$

Here, the number of classes is given as  $c$ ,  $x_{ij}$  ( $i, j = 1, 2, \dots, c$ ) is the number of illustrations of the  $i^{th}$  class is split obsessed by the  $j^{th}$  class. The  $x_{ii}$  elements on the diagonal stand in for the number of illustration arguments that remained impartially divided. The entire amount of illustration points is determined using  $n = \sum_{i=1}^c \sum_{j=1}^c x_{ij}$  where  $n$  represents the total number of values.

- Overall Accuracy (OA):** According to [307], the proportion of currently classified pixels to all pixels is referred to as overall accuracy

$$OA = \frac{1}{T} \sum_{c=1}^c T_{cc} \tag{92}$$

Here,  $T$  is the chosen classifier’s confusion matrix, and  $T_{cc}$  is the number of testing pixels.

- Average Accuracy (AA):** The average per-class classification accuracy is measured using average accuracy [307]. To determine the per-class proportion, we confidently calculate the ratio of pixels in a specific class to the total number of picture elements in that class.

$$AA = \frac{1}{C} \sum_{c=1}^c \frac{T_{cc}}{\sum_{c'=1}^c T_{cc'}} \tag{93}$$

Here,  $T$  is the number of challenging pixels, and  $T_{cc'}$  represents the confusion matrix of a given classifier.

- Kappa Coefficient (KC):** The Kappa coefficient is a measurement used for the number of tries to translate overall accuracy by dropping its worth when a promise might be gained via coincidental [307].

$$KC = \frac{\frac{1}{T} \sum_c T_{cc} - \frac{1}{T^2} (\sum_{c'} T_{cc'}) (\sum_{c'} T_{c'c})}{1 - \frac{1}{T^2} (\sum_{c'} T_{cc'}) (\sum_{c'} T_{c'c})} \tag{94}$$

- NDVI:** NDVI stands for Normalized Difference Vegetation Index. It is used to identify the probability of lower or higher vegetation. According to [307], the higher NDVI value indicates more vegetation cover, whereas the lower NDVI value indicates less vegetation cover. The NDVI readings fall within a range of -1 and +1.

$$NDVI = \frac{(NIR - RED)}{(NIR + RED)} \tag{95}$$

Here, NIR is characterized as the Near-Infrared Band, and the Red is denoted as the Red Band.

- NDWI:** NDWI stands for Normalized Difference Water Index. This describes the probability of having either low or high water content. According to [307], a higher NDWI number indicates a higher water content and a lower NDWI value indicates a lower water content. The NDWI values are in the range of -1 to +1.

$$NDWI = \frac{(NIR - SWIR)}{(NIR + SWIR)} \tag{96}$$

In this case, NIR represents the near-infrared band, and SWIR represents the short-wave infrared band.

7. **Rate and Percentage of Change:** Transformation rates and percentage changes are confidently and accurately calculated to demonstrate the precise LU/LC proportions for various time intervals.

$$POC = \left( \frac{T_2 - T_1}{T_1} \right) \times 100 \quad (97)$$

$$ROC \left( \frac{ha}{yr} \right) = \frac{T_1 - T_2}{T_i} \quad (98)$$

POC, or percentage of change, is used here. ROC is an acronym for rate of change. The variables  $T_1$  and  $T_2$  represent the area (ha) of LU/LC for time intervals 1 and 2, respectively, and  $T_i$  denotes the time interval in years (yr) between the two [308].

Other three categories, sensitivity, specificity, and accuracy, can be used to classify Hyperspectral Images. Sensitivity is used to identify the actual positives identified by the classifiers. The classifier uses specificity to determine the negatives identified as negatives. In the field of machine learning, a model's ability to accurately predict the class labels of previously unseen data is crucial. The overall accuracy of a model in this task is a key metric in assessing its performance.

$$Sensitivity = \frac{TP}{TP + FN} \quad (99)$$

$$Specificity = \frac{TN}{TN + FP} \quad (100)$$

$$Overall\ accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (101)$$

True positives, or TPs, are employed when circumstances are accurately identified, the test's results are positive, and the classification's actual value is positive. "False Positive" (FP) refers to an incorrectly discovered condition for which the test result was negative, but the classification was positive. The term "True Negative" (TN) refers to an accurately rejected condition. The test result is negative, but the classification's actual value is positive. The term "True Positives" (TP) refers to situations where a condition was wrongly rejected, the test result was positive, but the categorization had a negative value [309].

## 11 Open issues and challenges of hyperspectral image analysis

This section elaborates on the issues and challenges in the hyperspectral image analysis identified from the above study.

- Hyperspectral imaging involves capturing data across a wide range of spectral bands. However, not all bands contain useful information. Therefore, the process of identify-

ing the most relevant and informative bands is a complex and challenging task that requires careful analysis and interpretation of the data [102].

- Hyperspectral images, which capture information about objects at many different wavelengths across the electromagnetic spectrum, often present a challenge for image analysis algorithms due to the spatial complexity of objects in the image [115].
- Acquiring accurate and reliable labeled data is a crucial component in the development and training of machine learning models. However, the process of gathering labeled data can be a challenging and resource-intensive undertaking, particularly when dealing with large, complex, and diverse datasets [95].
- Achieving high classification accuracy rates and computational efficiency in hyperspectral imaging poses a significant challenge [55].
- Obtaining accurate classification of data requires the fusion of both spectral and spatial information. However, this can be a challenging task due to the complexity involved in combining these two types of information effectively [56].
- Preventing overfitting caused by multiple adjustable parameters is also a challenging task in hyperspectral imaging (HSI) [59].
- When pure spectra are used to classify hyperspectral imaging (HSI) data, the resulting low or medium spatial resolution may be caused by spectral mixture problems [77].
- One of the challenges in hyperspectral image classification is the automatic determination of the optimal number of superpixel segments [80].
- In hyperspectral imaging (HSI), misclassification between similar labels is a challenging problem that is difficult to overcome [110].

## 12 Discussion

According to Fig. 6, SVM, mixed convolution methods, and attention models give the best accuracy. When it comes to hyperspectral image classification, the SVM (Support Vector Machine) algorithm is a crucial tool that has proven its effectiveness time and time again. Its ability to handle high-dimensional data with many spectral bands makes it well-suited for dealing with hyperspectral images' complex and multidimensional nature. SVM is a powerful algorithm that effectively handles nonlinear decision boundaries commonly found in hyperspectral data. By finding the optimal hyperplane and maximizing the margin between different classes, SVM ensures that the classification results are accurate and reliable. Moreover, SVM's ability to generalize to new and unseen data makes it a reliable and robust approach for classification tasks. Overall, SVM's importance in hyperspectral image classification cannot be overstated, as it provides a powerful tool for researchers and practitioners to analyze and classify hyperspectral data accurately [92]. According to [119], CNNs are highly effective for hyperspectral image classification. They can handle high-dimensional data, learn features invariant to spectral variations, and classify these features into different classes. CNNs can also deal with the complex and nonlinear nature of hyperspectral data.

Convolutional Neural Networks (CNNs) have the remarkable ability to be trained with small amounts of labeled data, making them an invaluable tool for researchers and practitioners in the precise analysis and classification of hyperspectral data. The versatility of CNNs makes them a powerful approach for this purpose. Attention-based models have become increasingly popular in hyperspectral image classification as they can capture the

interdependencies of spectral bands while suppressing irrelevant information. These models assign weights to spectral bands based on their importance, allowing them to focus on the most relevant information. This results in improved performance with reduced computational complexity. Attention models provide an effective and efficient approach for accurately classifying hyperspectral data [109].

### 13 Conclusion and future directions

Hyperspectral image analysis is a complex process that entails multiple tasks, including pre-processing, feature extraction, band selection, classification, and prediction. This paper presents an in-depth review of several machine learning and deep learning approaches utilized in hyperspectral image analysis. Various classification techniques and their subcategories, including supervised, unsupervised, and deep learning-based, are also illustrated. Furthermore, the review covers the significant feature extraction methods specific to hyperspectral image analysis, such as spectral angle mapper, principal component analysis, and linear discriminant analysis. And a detailed description of the band selection techniques, including minimum noise fraction, principal component analysis, and successive projections algorithm. The paper discusses hyperspectral image analysis its significance, challenges, and real-world applications, including benchmark datasets and evaluation metrics. The review identifies the open issues and presents future directions that will aid researchers in effectively analyzing hyperspectral images.

Hyperspectral imaging is a critical field of study, especially in remote sensing and medical diagnosis. Despite the challenges posed by the high dimensionality of hyperspectral images, researchers and analysts continue to push the boundaries of what is possible. Their unwavering dedication and determination to overcome obstacles is truly inspiring. To address these challenges, future research directions are focused on developing advanced feature extraction and reduction techniques that can effectively reduce the dimensionality of data without compromising the quality of the results. Deep learning algorithms such as convolutional neural networks are a highly promising research area that efficiently extract features from high-dimensional data, in addition to feature extraction and reduction techniques. Researchers are also working on developing more efficient and scalable computing architectures such as parallel and distributed computing systems, to enable the processing and analysis of large hyperspectral datasets. One of the significant challenges of hyperspectral image analysis is band selection. Researchers are developing more advanced machine-learning algorithms to effectively identify and select relevant bands based on specific analysis requirements to address this challenge. Integrating domain knowledge and expert input into the analysis process can improve the accuracy and relevance of band selection.

Additionally, developing sophisticated visualization and exploration tools can help analysts better understand and interpret hyperspectral data, aiding in selecting appropriate bands. Multi-modal and multi-sensor data fusion techniques are also promising research areas for improving the accuracy and applicability of hyperspectral image analysis across various fields. To address the challenge of varying spectral signatures of objects in a hyperspectral image, researchers are developing advanced algorithms that can account for spectral variability and advanced machine learning techniques that can learn and adapt to the variability in data. Integrating multi-modal and multi-sensor data can also enhance the accuracy and reliability of hyperspectral image analysis in various applications. Another significant challenge in hyperspectral image analysis is the availability of labeled data for

training. To overcome this challenge, researchers are developing more efficient and practical techniques for data labeling, such as active learning and semi-supervised learning. Additionally, using transfer learning and pre-training on large datasets can reduce the labeled data needed for training. Furthermore, developing advanced unsupervised and weakly supervised learning algorithms can overcome the limitations of labeled data, making hyperspectral image analysis more accessible and applicable in various fields. Transformer architectures and attention mechanisms in deep learning models can be used to improve the classification accuracy of data. These models are capable of focusing on the most important elements of the input data and generating more precise and reliable predictions. By utilizing optimization algorithms, we can significantly improve the computational speed of various processes. These algorithms are designed to streamline operations by reducing the number of computations required to reach the desired outcome, resulting in faster and more efficient performance.

**Data availability** Data sharing does not apply to this article as no datasets were generated or analyzed during the current study.

## Declarations

**Conflict of interest** The authors declare that they have no conflict of interest.

## References

1. Adão T, Hruška J, Pádua L, Bessa J, Peres E, Morais R, Sousa JJ (2017) Hyperspectral imaging: A review on UAV-based sensors, data processing and applications for agriculture and forestry. *Remote Sens* 9:1110
2. Zaman Z, Ahmed SB, Malik MI (2023) Analysis of hyperspectral data to develop an approach for document images. *Sensors* 23:6845
3. Paoletti ME, Haut JM, Plaza J, Plaza A (2019) Deep learning classifiers for hyperspectral imaging: A review. *ISPRS J Photogramm Remote Sens* 158:279–317
4. Amigo JM, Babamoradi H, Elcoroaristizabal S (2015) Hyperspectral image analysis. A tutorial. *Anal Chim Acta* 896:34–51
5. Zhang M, Hu C, Kowalewski MG, Janz SJ (2017) Atmospheric correction of hyperspectral GCAS airborne measurements over the North Atlantic Ocean and Louisiana shelf. *IEEE Trans Geosci Remote Sens* 56:168–179
6. Guha A, Ghosh UK, Sinha J, Pour AB, Bhaisal R, Chatterjee S, Baranval NK, Rani N, Kumar KV, Rao PV (2021) Potentials of airborne hyperspectral AVIRIS-NG data in the exploration of base metal deposit—a study in the parts of Bhilwara, Rajasthan. *Remote Sens* 13:2101
7. Oppelt N, Mauser W (2007) Airborne visible/infrared imaging spectrometer AVIS: Design, characterization, and calibration. *Sensors* 7:1934–1953
8. Noor NRM, Vladimirova T (2013) Investigation into lossless hyperspectral image compression for satellite remote sensing. *Int J Remote Sens* 34:5072–5104
9. Morales G, Sheppard JW, Logan RD, Shaw JA (2021) Hyperspectral dimensionality reduction based on inter-band redundancy analysis and greedy spectral selection. *Remote Sens* 13:3649
10. Aasen H, Honkavaara E, Lucieer A, Zarco-Tejada PJ (2018) Quantitative remote sensing at ultra-high resolution with UAV spectroscopy: a review of sensor technology, measurement procedures, and data correction workflows. *Remote Sens* 10:1091
11. Rogers M, Blanc-Talon J, Urschler M, Delmas P (2023) Wavelength and texture feature selection for hyperspectral imaging: a systematic literature review. *J Food Meas Charact* 17:6039–6064
12. Chang CY, Zhou R, Kira O, Marri S, Skovira J, Gu L, Sun Y (2020) An Unmanned Aerial System (UAS) for concurrent measurements of solar induced chlorophyll fluorescence and hyperspectral reflectance toward improving crop monitoring. *Agric Forest Meteorol* 294:108145

13. Shippert P (2004) Why use hyperspectral imagery? *Photogramm Eng Remote Sens* 70:377–396
14. Awan R, Al-Maadeed S, Al-Saady R (2018) Using spectral imaging for the analysis of abnormalities for colorectal cancer: When is it helpful? *PLoS One* 13:e0197431
15. Purnamasayangasukasih PR, Norizah K, Ismail AA, Shamsudin I (2016) A review of uses of satellite imagery in monitoring mangrove forests. *IOP Conf Ser Earth Environ Sci* 37:012034
16. Reguzzoni M, Sanso F, Venuti G, Brivio PA (2003) Bayesian classification by data augmentation. *Int J Remote Sens* 24:3961–3981
17. Kruse FA, Taranić JV, Coolbaugh M, Michaels J, Littlefield EF, Calvin WM, Martini BA (2011) Effect of reduced spatial resolution on mineral mapping using imaging spectrometry—Examples using Hyperspectral Infrared Imager (HyspIRI) simulated data. *Remote Sensing* 3:1584–1602
18. Kumar MS, Keerthi V, Anjain RN, Sarma MM, Bothale V (2020) Evaluation of machine learning methods for hyperspectral image classification. In: 2020 IEEE India geoscience and remote sensing symposium (inGARSS), pp 225–228
19. Yang M, Hu Y, Tian H, Khan FA, Liu Q, Goes JI, Kim W (2021) Atmospheric correction of airborne hyperspectral CASI data using polymer, 6S and FLAASH. *Remote Sens* 13:5062
20. Zhu L, Wen G, Qiu S (2018) Low-rank and sparse matrix decomposition with cluster weighting for hyperspectral anomaly detection. *Remote Sens* 10:707
21. Horig B, Kühn F, Oschütz F, Lehmann F (2001) HyMap hyperspectral remote sensing to detect hydrocarbons. *Int J Remote Sens* 22:1413–1422
22. Cheng Y-B, Ustin SL, Riaño D, Vanderbilt VC (2008) Water content estimation from hyperspectral images and MODIS indexes in Southeastern Arizona. *Remote Sens Environ* 112:363–374
23. Ma L, Crawford MM, Tian J (2010) Local manifold learning-based  $k$   $k$ -nearest-neighbor for hyperspectral image classification. *IEEE Trans Geosci Remote Sens* 48:4099–4109
24. Cutter MA (2004) Compact high-resolution imaging spectrometer (CHRIS) design and performance. *Imaging Spectrometry X* 5546:126–131
25. Zhong Y, Hu X, Luo C, Wang X, Zhao J, Zhang L (2020) WHU-Hi: UAV-borne hyperspectral with high spatial resolution (H2) benchmark datasets and classifier for precise crop identification based on deep convolutional neural network with CRF. *Remote Sens Environ* 250:112012
26. Saita Y, Shimoyama D, Takahashi R, Nomura T (2022) Single-shot compressive hyperspectral imaging with dispersed and undispersed light using a generally available grating. *Appl Opt* 61:1106
27. [https://sensing.konicaminolta.asia/hyperspectral-imaging-solutions/?utm\\_term=hyperspectral%20imaging&utm\\_campaign=\(All\)+SSG2021+-+Specim&utm\\_source=adwords&utm\\_medium=ppc&hsa\\_acc=5702629821&hsa\\_cam=13951520117&hsa\\_grp=127893723547&hsa\\_ad=534418255585&hsa.](https://sensing.konicaminolta.asia/hyperspectral-imaging-solutions/?utm_term=hyperspectral%20imaging&utm_campaign=(All)+SSG2021+-+Specim&utm_source=adwords&utm_medium=ppc&hsa_acc=5702629821&hsa_cam=13951520117&hsa_grp=127893723547&hsa_ad=534418255585&hsa.) [Online]
28. Raj R, Walker JP, Vinod V, Pingale R, Naik B, Jagarlapudi A (2021) Leaf water content estimation using top-of canopy airborne hyperspectral data. *Int J Appl Earth Obs Geoinformation* 102:102393
29. Lin Z, Chen Y, Zhao X, Wang G (2013) Spectral-spatial classification of hyperspectral image using autoencoders. In: 9th international conference on information, communications & signal processing
30. Khan MJ, Khan HS, Yousaf A, Khurshid K, Abbas A (2018) Modern trends in hyperspectral image analysis: A review. *IEEE Access* 6:14118–14129
31. Dong Z, Cai Y, Cai Z, Liu X, Yang Z, Zhuge M (2020) Cooperative spectral–spatial attention dense network for hyperspectral image classification. *IEEE Geosci Remote Sens Lett* 18:866–870
32. Huo H, Guo J, Li Z-L (2018) Hyperspectral image classification for land cover based on an improved interval type-II fuzzy C-means approach. *Sensors* 18:363
33. Alshari EA, Gawali BW (2021) Development of classification system for LULC using remote sensing and GIS. *Glob Transit proc* 2:8–17
34. Saha D, Manickavasagan A (2021) Machine learning techniques for analysis of hyperspectral images to determine quality of food products: A review. *Curr Res Food Sci* 4:28–44
35. Signoroni A, Savardi M, Baronio A, Benini S (2019) Deep learning meets hyperspectral image analysis: A multidisciplinary review. *J Imaging* 5:52
36. Datta D, Mallick PK, Bhoi AK, Ijaz MF, Shafi J, Choi J (2022) Hyperspectral image classification: Potentials, challenges, and future directions. *Comput Intell Neurosci* 2022:1–36
37. Bera S, Shrivastava VK, Satapathy SC (2022) Advances in hyperspectral image classification based on convolutional neural networks: A review. *CMES-Comput Model Eng Sci* 133:219–250
38. Kaur G, Saini KS, Singh D, Kaur M (2021) A comprehensive study on computational pansharpening techniques for remote sensing images. *Arch Comput Methods Eng* 28:4961–4978
39. Singh D, Kaur M, Singh H (2018) Remote sensing image fusion using fuzzy logic and gyrator transform. *Remote Sens Lett* 9:942–951

40. Singh D, Kaur M, Jabarulla MY, Kumar V, Lee H-N (2022) Evolving fusion-based visibility restoration model for hazy remote sensing images using dynamic differential evolution. *IEEE Trans Geosci Remote Sens* 60:1–14
41. Zhao J, Yan H, Huang L (2023) A joint method of spatial–spectral features and BP neural network for hyperspectral image classification. *Egypt J Remote Sens Space Sci* 26:107–115
42. Yao D, Zhi-li Z, Xiao-feng Z, Wei C, Fang H, Yao-ming C, Cai W-W (2023) Deep hybrid: multi-graph neural network collaboration for hyperspectral image classification. *Defence Technol* 23:164–176
43. Soltani A, El Mannai H, Sabeur MN (2023) A new hyperspectral image classification method based on extended wavelets transform. In: 2023 IEEE international conference on advanced systems and emergent technologies, pp 1–5
44. Rujan L, Neagoe VE (2023) A novel approach for hyperspectral image classification using bat algorithm to optimize a CNN classifier. In: 2023 15th international conference on electronics, computers and artificial intelligence (ECAI), pp 1–6
45. Xiong X, Huang Z, Sheng Z, Xu Z, He X, Lin Q (2022) A three-dimensional cyclic multiplication-convolution neural network for hyperspectral image classification. In: 2022 IEEE Smartworld, ubiquitous intelligence & computing, scalable computing & communications, digital twin, privacy computing, Metaverse, autonomous & trusted vehicles (SmartWorld/UIC/ScalCom/DigitalTwin/PriComp/Meta), pp 2506–2512
46. Li S, Zhang Q, Cheng L, Peng B (2023) A two stage learning algorithm for hyperspectral image classification. In: 2023 5th international conference on natural language processing (ICNLP), pp 86–91
47. Bellio G, Russell R, Kursun O (2023) Boosting with multiple clustering memberships for hyperspectral image classification. *SoutheastCon* 2023:175–178
48. Liao D, Shi C, Wang L (2023) A complementary integrated transformer network for hyperspectral image classification. *CAAI Trans Intell Technol* 8:1288–1307
49. Sunkara R, Singh AK, Kadambi GR (2023) Class information-based principal component analysis algorithm for improved hyperspectral image classification. In: 2023 international conference on machine intelligence for GeoAnalytics and remote sensing (MIGARS), vol 1, pp 1–4
50. Lei Y, Zhao G, Zhang L (2023) Deep composite kernels ELM based on spatial feature extraction for hyperspectral vegetation image classification. In: 5th international conference on natural language processing (ICNLP)
51. Li H-C, Lin Z-X, Ma T-Y, Zhao X-L, Plaza A, Emery WJ (2023) Hybrid fully connected tensorized compression network for hyperspectral image classification. *IEEE Trans Geosci Remote Sens* 61:1–16
52. Wei L, Ma H, Yin Y (2023) Improved K-means Hyperspectral Image Classification Algorithm Based on Variance Coefficient Weighting. In 2023 3rd Asia-Pacific Conference on Communications Technology and Computer Science (ACCTCS) 497–501
53. Dündar T, İnce T (2023) Locality constraint joint-sparse and weighted low-rank based hyperspectral image classification. In: 2023 10th international conference on recent advances in air and space technologies (RAST), pp 1–6
54. Scheibenreif L, Mommert M, Borth D (2023) Masked vision transformers for hyperspectral image classification. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 2165–2175
55. Chen GY (2021) Multiscale filter-based hyperspectral image classification with PCA and SVM. *J Electr Eng* 72:40–45
56. Zhang S, Huang H, Huang Y, Cheng D, Huang J (2022) A GA and SVM classification model for pine wilt disease detection using UAV-based hyperspectral imagery. *Appl Sci* 12:6676
57. Pathak DK, Kalita SK, Bhattacharya DK (2021) Hyperspectral image classification using support vector machine: a spectral spatial feature based approach. *Evol Intell* 15:1809–1823
58. Shang Y, Zheng X, Li J, Liu D, Wang P (2022) A comparative analysis of swarm intelligence and evolutionary algorithms for feature selection in SVM-based hyperspectral image classification. *Remote Sens* 14:3019
59. Guo B, Gunn SR, Damper RI, Nelson JD (2008) Customizing kernel functions for SVM-based hyperspectral image classification. *IEEE Trans Image Process* 17:622–629
60. Ren J, Wang R, Liu G, Wang Y, Wu W (2020) An SVM-based nested sliding window approach for spectral–spatial classification of hyperspectral images. *Remote Sens* 13:114
61. Zhang N, Wang Y, Zhang X (2020) Extraction of tree crowns damaged by *Dendrolimus tabulaeformis* Tsai et Liu via spectral-spatial classification using UAV-based hyperspectral images. *Plant Methods* 16:1–19



62. Champa AI, Rabbi MF, Hasan SM, Zaman A, Kabir MH (2021) Tree-based classifier for hyperspectral image classification via hybrid technique of feature reduction. In: International conference on information and communication technology for sustainable development (ICICT4SD)
63. Xu S, Liu S, Wang H, Chen W, Zhang F, Xiao Z (2020) A hyperspectral image classification approach based on feature fusion and multi-layered gradient boosting decision trees. *Entropy* 23:20
64. Wei L, Huang C, Wang Z, Wang Z, Zhou X, Cao L (2019) Monitoring of urban black-odor water based on Nemerow index and gradient boosting decision tree regression using UAV-borne hyperspectral imagery. *Remote Sens* 11:2402
65. Qasim M, Khan SD (2022) Detection and relative quantification of neodymium in sillai patti carbonatite using decision tree classification of the hyperspectral data. *Sensors* 22:7537
66. Wang L, Wang Q (2022) Fast spatial-spectral random forests for thick cloud removal of hyperspectral images. *Int J Appl Earth Obs Geoinformation* 112:102916
67. Jain V, Phophalia A (2019) Exponential weighted random forest for hyperspectral image classification. In: IGARSS 2019-2019 IEEE international geoscience and remote sensing symposium, pp 3297–3300
68. Kishore KMS, Behera MK, Chakravarty S, Dash S (2020) Hyperspectral image classification using minimum noise fraction and random forest. In: IEEE international women in engineering (WIE) conference on electrical and computer engineering (WIECON-ECE)
69. Wang A, Wang Y, Chen Y (2019) Hyperspectral image classification based on convolutional neural network and random forest. *Remote Sens Lett* 10:1086–1094
70. Liu B, Guo W, Chen X, Gao K, Zuo X, Wang R, Yu A (2020) Morphological attribute profile cube and deep random forest for small sample classification of hyperspectral image. *IEEE Access* 8:117096–117108
71. Bazine R, Huayi W, Boukhechba K (2019) K-NN similarity measure based on fourier descriptors for hyperspectral images classification. In: Proceedings of the 2019 international conference on video, signal and image processing, pp 39–43
72. Bhavatarini N, Akash BN, Avinash AR, Akshay HM (2023) Object detection and classification of hyperspectral images using K-NN. In: 2023 Second International Conference on Electrical, Electronics, Information and Communication Technologies (ICEEICT), pp 1–6
73. Moiane A, Machado AML (2019) Class-based affinity propagation for hyperspectral image dimensionality reduction and improvement of maximum likelihood classification accuracy. *Boletim de Ciências Geodésicas* 25
74. Liu N, Guo Y, Jiang H, Yi W (2020) Gastric cancer diagnosis using hyperspectral imaging with principal component analysis and spectral angle mapper. *J Biomed Opt* 25:1
75. Christovam LE, Pessoa GG, Shimabukuro MH, Galo MLBT (2019) Land use and land cover classification using hyperspectral imagery: Evaluating the performance of spectral angle mapper, support vector machine and random forest. *The international archives of the photogrammetry, remote sensing and spatial information sciences* 42:1841–1847
76. Wei L, Ma H, Yin Y, Geng C (2023) Kmeans-CM algorithm with spectral angle mapper for hyperspectral image classification. *IEEE Access* 11:26566–26576
77. Ahmad M, Khan A, Khan AM, Mazzara M, Distefano S, Sohaib A, Nibouche O (2019) Spatial prior fuzziness pool based interactive classification of hyperspectral images. *Remote Sens* 11:1136
78. Ma A, Filippi AM, Wang Z, Yin Z (2019) Hyperspectral image classification using similarity measurements-based deep recurrent neural networks. *Remote Sens* 11:194
79. Wong M, Abeysinghe W, Hung CC (2019) A massive self-organizing map for hyperspectral image classification. In: 2019 10th Workshop on hyperspectral imaging and signal processing: Evolution in remote sensing (WHISPERS), pp 1–5
80. Zhang X, Jiang X, Jiang J, Zhang Y, Liu X, Cai Z (2021) Spectral-spatial and superpixelwise PCA for unsupervised feature extraction of hyperspectral imagery. *IEEE Trans Geosci Remote Sens* 60:1–10
81. Zhang M, Gong M, He H, Zhu S (2020) Symmetric all convolutional neural-network-based unsupervised feature extraction for hyperspectral images classification. *IEEE Trans Cybern* 52:2981–2993
82. Nalepa J, Myller M, Imai Y, Honda K-I, Takeda T, Antoniuk M (2020) Unsupervised segmentation of hyperspectral images using 3-D convolutional autoencoders. *IEEE Geosci Remote Sens Lett* 17:1948–1952
83. Liu S, Chu RS, Wang X, Luk W (2019) Optimizing CNN-based hyperspectral image classification on FPGAs. In: International symposium on applied reconfigurable computing, pp 17–31
84. Zhang L, Huang D, Chen X, Zhu L, Xie Z, Chen X, Shi W (2023) Discrimination between normal and necrotic small intestinal tissue using hyperspectral imaging and unsupervised classification. *J Biophotonics* e202300020

85. Shi G, Huang H, Wang L (2019) Unsupervised dimensionality reduction for hyperspectral imagery via local geometric structure feature learning. *IEEE Geosci Remote Sens Lett* 17:1425–1429
86. Xu X, Li J, Li S, Plaza A (2019) Subpixel component analysis for hyperspectral image classification. *IEEE Trans Geosci Remote Sens* 57:5564–5579
87. Su H, Yu Y, Wu Z, Du Q (2020) Random subspace-based k-nearest class collaborative representation for hyperspectral image classification. *IEEE Trans Geosci Remote Sens* 59:6840–6853
88. Chen H, Miao F, Chen Y, Xiong Y, Chen T (2021) A hyperspectral image classification method using multifeature vectors and optimized KELM. *IEEE J Sel Top Appl Earth Obs Remote Sens* 14:2781–2795
89. Gao H, Yang Y, Li C, Gao L, Zhang B (2020) Multiscale residual network with mixed depthwise convolution for hyperspectral image classification. *IEEE Trans Geoscience Remote Sens* 59:3396–3408
90. Zhang Y, Cao G, Li X (2020) Multiview-based random rotation ensemble pruning for hyperspectral image classification. *IEEE Trans Instrum Meas* 70:1–14
91. Camps-Valls G, Marsheva TVB, Zhou D (2007) Semi-supervised graph-based hyperspectral image classification. *IEEE Trans Geosci Remote Sens* 45:3044–3054
92. Manian V, Alfaro-Mejía E, Tokars RP (2022) Hyperspectral image labeling and classification using an ensemble semi supervised machine learning approach. *Sensors* 22:1623
93. Zeng H, Liu Q, Zhang M, Han X, Wang Y (2020) Semi-supervised hyperspectral image classification with graph clustering convolutional networks. *arXiv preprint arXiv:2012.10932*
94. Zheng X, Jia J, Chen J, Guo S, Sun L, Zhou C, Wang Y (2022) Hyperspectral image classification with imbalanced data based on semi-supervised learning. *Appl Sci* 12:3943
95. Liu L, Hong D, Ni L, Gao L (2022) Multilayer cascade screening strategy for semi-supervised change detection in hyperspectral images. *IEEE J Sel Top Appl Earth Obs Remote Sens* 15:1926–1940
96. Wu Y, Mu G, Qin C, Miao Q, Ma W, Zhang X (2020) Semi-supervised hyperspectral image classification via spatial regulated self-training. *Remote Sensing* 12:159
97. Lei Z, Yi Z, Peng L, Hui SX (2020) Semi-supervised classification of hyperspectral images based on two branch autoencoder. *IOP Conf Ser Earth Environ Sci* 502:012014
98. Feng Z, Yang S, Wang M, Jiao L (2019) Learning dual geometric low-rank structure for semisupervised hyperspectral image classification. *IEEE Trans Cybern* 51(1):346–358
99. Hu W-S, Li H-C, Deng Y-J, Sun X, Du Q, Plaza A (2021) Lightweight tensor attention-driven ConvL-STM neural network for hyperspectral image classification. *IEEE J Sel Top Signal Process* 15:734–745
100. Hu L, Qi C, Wang Q (2018) Spectral-spatial hyperspectral image classification based on mathematical morphology post-processing. *Proc Comput Sci* 129:93–97
101. Wang J, Gao F, Dong J, Du Q (2020) Adaptive dropblock-enhanced generative adversarial networks for hyperspectral image classification. *IEEE Trans Geosci Remote Sens* 59:5040–5053
102. Samat A, Li E, Du P, Liu S, Xia J (2021) GPU-accelerated CatBoost-forest for hyperspectral image classification via parallelized mRMR ensemble subspace feature selection. *IEEE J Sel Top Appl Earth Obs Remote Sens* 14:3200–3214
103. Tu B, Zhou C, Liao X, Zhang G, Peng Y (2020) Spectral-spatial hyperspectral classification via structural-kernel collaborative representation. *IEEE Geosci Remote Sens Lett* 18:861–865
104. Roy SK, Manna S, Song T, Bruzzone L (2020) Attention-based adaptive spectral-spatial kernel ResNet for hyperspectral image classification. *IEEE Trans Geosci Remote Sens* 59:7831–7843
105. Lu Z, Xu B, Sun L, Zhan T, Tang S (2020) 3-D channel and spatial attention based multiscale spatial-spectral residual network for hyperspectral image classification. *IEEE J Sel Top Appl Earth Obs Remote Sens* 13:4311–4324
106. Wang L, Wang L, Wang Q, Atkinson PM (2021) SSA-SiamNet: Spectral-spatial-wise attention-based Siamese network for hyperspectral image change detection. *IEEE Trans Geosci Remote Sens* 60:1–18
107. Xu Q, Yuan X, Ouyang C, Zeng Y (2020) Attention-based pyramid network for segmentation and classification of high-resolution and hyperspectral remote sensing images. *Remote Sens* 12:3501
108. Mdrafi R, Du Q, Gurbuz AC, Tang B, Ma L, Younan NH (2020) Attention-based domain adaptation using residual network for hyperspectral image classification. *IEEE J Sel Top Appl Earth Obs Remote Sens* 13:6424–6433
109. AL-Kubaisi MA, Shafri HZ, Ismail MH, MJM Y, Jahari bin Hashim S (2023) Attention-based multi-scale deep learning with unsampled pixel utilization for hyperspectral image classification. *Geocarto Int* 38:2231428
110. Hsieh T-H, Kiang J-F (2020) Comparison of CNN algorithms on hyperspectral image classification in agricultural lands. *Sensors* 20:1734
111. Yu C, Han R, Song M, Liu C, Chang C-I (2020) A simplified 2D–3D CNN architecture for hyperspectral image classification based on spatial-spectral fusion. *IEEE J Sel Top Appl Earth Obs Remote Sens* 13:2485–2501

112. Ghaderizadeh S, Abbasi-Moghadam D, Sharifi A, Zhao N, Tariq A (2021) Hyperspectral image classification using a hybrid 3D–2D convolutional neural networks. *IEEE J Sel Top Appl Earth Obs Remote Sens* 14:7570–7588
113. Liu J, Zhang K, Wu S, Shi H, Zhao Y, Sun Y, Zhuang H, Fu E (2022) An investigation of a multidimensional CNN combined with an attention mechanism model to resolve small-sample problems in hyperspectral image classification. *Remote Sens* 14:785
114. Hang R, Li Z, Liu Q, Ghamisi P, Bhattacharyya SS (2020) Hyperspectral image classification with attention-aided CNNs. *IEEE Trans Geosci Remote Sens* 59:2281–2293
115. Diakite A, Jiangsheng G, Xiaping F (2021) Hyperspectral image classification using 3D 2D CNN. *IET Image Process* 15:1083–1092
116. Dong S, Quan Y, Feng W, Dauphin G, Gao L, Xing M (2021) A pixel cluster CNN and spectral-spatial fusion algorithm for hyperspectral image classification with small-size training samples. *IEEE J Sel Top Appl Earth Obs Remote Sens* 14:4101–4114
117. Ge Z, Cao G, Li X, Fu P (2020) Hyperspectral image classification method based on 2D–3D CNN and multibranch feature fusion. *IEEE J Sel Top Appl Earth Obs Remote Sens* 13:5776–5788
118. Butt MHF, Ayaz H, Ahmad M, Li JP, Kuleev R (2022) A fast and compact hybrid CNN for hyperspectral imaging based bloodstain classification. In: 2022 IEEE Congress on Evolutionary Computation (CEC), pp 1–8
119. Zheng J, Feng Y, Bai C, Zhang J (2020) Hyperspectral image classification using mixed convolutions and covariance pooling. *IEEE Trans Geosci Remote Sens* 59
120. Bhosle K, Ahirwadkar B (2021) Deep learning convolutional neural network (CNN) for cotton, mulberry and sugarcane classification using hyperspectral remote sensing data. *J IntSci Technol* 9:70–74
121. Yan T, Xu W, Lin J, Duan L, Gao P, Zhang C, Lv X (2021) Combining multi-dimensional convolutional neural network (CNN) with visualization method for detection of aphid gossypii glover infection in cotton leaves using hyperspectral imaging. *Front Plant Sci* 12 hyperspectral imaging techniques. *IEEE Access* 8:123026–123036
122. Pang L, Men S, Yan L, Xiao J (2020) Rapid vitality estimation and prediction of corn seeds based on spectra and images using deep learning and hyperspectral imaging techniques. *Ieee Access* 8:123026–123036
123. Kjør A, Nielsen G, Størke S, Clausen MR, Edelenbos M, Jørgensen B (2016) Prediction of starch, soluble sugars and amino acids in potatoes (*Solanum tuberosum* L.) using hyperspectral imaging, dielectric and LF-NMR methodologies. *Potato Res* 59:357–374
124. Wang SW, Munkhnasan L, Lee W-K (2021) Land use and land cover change detection and prediction in Bhutan's high altitude city of Thimphu, using cellular automata and Markov chain. *Environ Challenges* 2:100017
125. Nguyen TTH, Ngo TTP (2018) Land use/land cover change prediction in Dak Nong Province based on remote sensing and Markov Chain Model and Cellular Automata. *J Viet Environ* 9:132–140
126. Nguyen HTT, Pham TA, Doan MT, Tran PTX (2020) Land use/land cover change prediction using multi-temporal satellite imagery and multi-layer perceptron Markov model. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 44:99–105
127. Vinayak B, Lee HS, Gedem S (2021) Prediction of land use and land cover changes in Mumbai City, India, using remote sensing data and a multilayer perceptron neural network-based Markov chain model. *Sustainability* 13:471
128. Cunha ERD, Santos CAG, Silva RMD, Bacani VM, Pott A (2021) Future scenarios based on a CA-Markov land use and land cover simulation model for a tropical humid basin in the Cerrado/Atlantic forest ecotone of Brazil. *Land Use Policy* 101:105141
129. Pang L, Wang J, Men S, Yan L, Xiao J (2021) Hyperspectral imaging coupled with multivariate methods for seed vitality estimation and forecast for *Quercus variabilis*. *Spectrochim Acta Part A: Mol Biomol Spectrosc* 245:118888
130. Jiang X, Zhen J, Miao J, Zhao D, Wang J, Jia S (2021) Assessing mangrove leaf traits under different pest and disease severity with hyperspectral imaging spectroscopy. *Ecol Indicators* 129:107901
131. Boggavarapu LPK, Manoharan P (2020A) new framework for hyperspectral image classification using Gabor embedded patch based convolution neural network. *Infrared Phys Technol* 110:103455
132. Huang M, Tang J, Yang B, Zhu Q (2016) of maize seeds of different years based on hyperspectral imaging and model updating. *Comput Electr Agric* 122:139–145
133. Zhang T, Wei W, Zhao B, Wang R, Li M, Yang L, Wang J, Sun Q (2018) A reliable methodology for determining seed viability by using hyperspectral data from two sides of wheat seeds. *Sensors* 18:813
134. Ambrose A, Kandpal LM, Kim MS, Lee W-H, Cho B-K (2016) High speed measurement of corn seed viability using hyperspectral imaging. *Infrared Phys Technol* 75:173–179
135. Baianu I (2011) Applications of microspectroscopy, hyperspectral chemical imaging and fluorescence microscopy in chemistry, biochemistry, biotechnology, molecular and cell biology. *Nature Prece*:1–1

136. Venkatesh R, Reddy NG, Pulipaka RS, Pereira A (2021) Rare presentation of choroidal neovascularisation in a case of congenital hypertrophy of retinal pigment epithelium. *BMJ Case Reports CP* 14:e244554
137. Lan W, Jaillais B, Renard CM, Leca A, Chen S, Bourvellec CL, Bureau S (2021) A method using near infrared hyperspectral imaging to highlight the internal quality of apple fruit slices. *Postharvest Biol Technol* 175:111497
138. Stuart MB, Davies M, Hobbs MJ, Pering TD, McGonigle AJ, Willmott JR (2022) High-resolution hyperspectral imaging using low-cost components: Application within environmental monitoring scenarios. *Sensors* 22:4652
139. Smith KM, Larive LL, Romanelli F (2002) Club drugs: methylenedioxymethamphetamine, flunitrazepam, ketamine hydrochloride, and gammahydroxybutyrate. *Am J Health-Syst Pharm* 59:1067–1076
140. Devassy BM, Georg S (2021) Forensic analysis of beverage stains using hyperspectral imaging. *Sci Rep* 11
141. Almeida MR, Logrado LPL, Zacca JJ, Correa DN, Poppi RJ (2017) Raman hyperspectral imaging in conjunction with independent component analysis as a forensic tool for explosive analysis: The case of an ATM explosion. *Talanta* 174:628–632
142. Chandran Suja V, Sentmanat J, Hofmann G, Scales C, Fuller GG (2020) Hyperspectral imaging for dynamic thin film interferometry. *Sci Rep* 10:11378
143. Favreau P, Hernandez C, Lindsey AS, Alvarez DF, Rich T, Prabhat P, Leavesley SJ (2014) Thin-film tunable filters for hyperspectral fluorescence microscopy. *J Biomed Opt* 19:011017–011017
144. Kuska MT, Behmann J, Mahlein A-K (2018) Potential of hyperspectral imaging to detect and identify the impact of chemical warfare compounds on plant tissue. *Pure Appl Chem* 90:1615–1624
145. Lu G, Fei B (2014) Medical hyperspectral imaging: a review. *J Biomed Opt* 19:010901
146. Kulcke A, Holmer A, Wahl P, Siemers F, Wild T, Daeschlein G (2018) A compact hyperspectral camera for measurement of perfusion parameters in medicine. *Biomedical Engineering/Biomedizinische Technik* 63:519–527
147. Meena SD, Agilandeeswari L (2019) An efficient framework for animal breeds classification using semi supervised learning and multi-part convolutional neural network (MP-CNN). *IEEE Access* 7:151783–151802
148. Meena SD, Loganathan A (2020) Intelligent animal detection system using sparse multi discriminative-neural network (SMD-NN) to mitigate animalvehicle collision. *Environ Sci Pollut Res* 27:39619–39634
149. Meena D, Agilandeeswari L (2020) Invariant features-based fuzzy inference system for animal detection and recognition using thermal images. *Int J Fuzzy Syst* 22:1868–1879
150. Sundaram DM, Loganathan A (2020) FSSCaps-DetCountNet: fuzzy soft sets and CapsNet-based detection and counting network for monitoring animals from aerial images. *J Appl Remote Sens* 14:1
151. Meena SD, Agilandeeswari L (2021) Smart animal detection and counting framework for monitoring livestock in an autonomous unmanned ground vehicle using restricted supervised learning and image fusion. *Neural Process Lett* 53:1253–1285
152. Frouin RJ, Franz BA, Ibrahim A, Knobelspiess K, Ahmad Z, Cairns B, Zhai PW (2019) Atmospheric correction of satellite ocean-color imagery during the PACE era. *Front Earth Sci* 7:145
153. Mao Z, Tao B, Chen J, Chen P, Hao Z, Zhu Q, Huang H (2020) A layer removal scheme for atmospheric correction of satellite ocean color data in coastal regions. *IEEE Trans Geosci Remote Sens* 59:1382–1391
154. Wright LA, Kindel BC, Pilewskie P, Leisso NP, Kampe TU, Schmidt KS (2020) Below-cloud atmospheric correction of airborne hyperspectral imagery using simultaneous solar spectral irradiance observations. *IEEE Trans Geosci Remote Sens* 59:1392–1409
155. Simard M, Riel BV, Denbina M, Hensley S (2016) Radiometric correction of airborne radar images over forested terrain with topography. *IEEE Trans Geosci Remote Sens* 54:4488–4500
156. Luo S, Tong L (2020) Radiometric correction of dual-polarization SAR data over steep terrain. In: *IGARSS 2020-2020 IEEE international geoscience and remote sensing symposium*, pp 1552–1555
157. Duan Y, Chen W, Wang M, Yan L (2013) A relative radiometric correction method for airborne image using outdoor calibration and image statistics. *IEEE Trans Geosci Remote Sens* 52:5164–5174
158. Goswami A, Mathuku DSH, Gangadharan SMP, Yadav CS, Sahu SK, Pradhan MK, Singh J, Imran H (2022) Change detection in remote sensing image data comparing algebraic and machine learning methods. *Electronics* 11:431
159. Green EP, Clark CD, Edwards AJ (2000) Geometric correction of satellite and airborne imagery. In: *Remote sensing handbook for tropical coastal management. Coastal management sourcebooks 3*, UNESCO, Paris, pp 93–108
160. Ahmad M, Shabbir S, Raza RA, Mazzara M, Distefano S, Khan AM (2021) Artifacts of different dimension reduction methods on hybrid CNN feature hierarchy for hyperspectral image classification. *Optik* 246:167757

161. Sawant S, Manoharan P (2020) Hyperspectral band selection based on metaheuristic optimization approach. *Infrared Phys Technol* 107:103295
162. Chang J, Chang C-I (2006) Independent component analysis-based dimensionality reduction with applications in hyperspectral image analysis. *IEEE Trans Geosci Remote Sens* 44:1586–1600
163. Devassy BM, George S (2020) Dimensionality reduction and visualisation of hyperspectral ink data using t-SNE. *Forensic Sci Int* 311:110194
164. Florimbi G, Fabelo H, Torti E, Ortega S, Marrero-Martin M, Callico GM, Danese G, Leporati F (2020) Towards real time computing of intraoperative hyperspectral imaging for brain cancer detection using multi-GPU platforms. *IEEE Access* 8:8485–8501
165. Fabelo H, Ortega S, Guerra R, Callicó G, Szolna A, Piñeiro JF, Sarmiento R (2016) A novel use of hyperspectral images for human brain cancer detection using in-vivo samples. In: *Special session on smart embedded biomedical devices for in situ physiological signal processing*, vol 5, pp 311–320
166. Ruiz L, Martín A, Urbanos G, Villanueva M, Sancho J, Rosa G, Sanz C (2020) Multiclass brain tumor classification using hyperspectral imaging and supervised machine learning. In: *2020 XXXV conference on design of circuits and integrated systems (DCIS)*, pp 1–6
167. Li J, Li X, Yan Y (2023) Unlocking the potential of data augmentation in contrastive learning for hyperspectral image classification. *Remote Sens* 15:3123
168. Liu N, Li W, Du Q (2018) Unsupervised feature extraction for hyperspectral imagery using collaboration competition graph. *IEEE J Sel Top Signal Process* 12:1491–1503
169. Li L, Ge H, Gao J, Zhang Y, Tong Y, Sun J (2020) A novel geometric mean feature space discriminant analysis method for hyperspectral image feature extraction. *Neural Process Lett* 51:515–542
170. Jia W, Sun M, Lian J, Hou S (2022) Feature dimensionality reduction: a review. *Complex Intell Syst* 8:2663–2693
171. Gisbrecht A, Schulz A, Hammer B (2015) Parametric nonlinear dimensionality reduction using kernel t-SNE. *Neurocomputing* 147:71–82
172. Palani A, Loganathan A (2024) Semi-Blind watermarking using convolutional attention-based turtle shell matrix for tamper detection and recovery of medical images. *Exp Syst Appl* 238:121903
173. Gao X, Sun Q, Xu H, Li Y (2018) 2D-LPCCA and 2D-SPCCA: Two new canonical correlation methods for feature extraction, fusion, and recognition. *Neurocomputing* 284:148–159
174. Choi SW, Lee C, Lee J-M, Park JH, Lee I-B (2005) Fault detection and identification of nonlinear processes based on kernel PCA. *Chemometrics Intell Lab Syst* 75:55–67
175. Kuo B-C, Landgrebe DA (2004) Nonparametric weighted feature extraction for classification. *IEEE Trans Geosci Remote Sens* 42:1096–1105
176. Rasti B, Hong D, Hang R, Ghamisi P, Kang X, Chanussot J, Benediktsson JA (2020) Feature extraction for hyperspectral imagery: The evolution from shallow to deep: Overview and toolbox. *IEEE Geosci Remote Sens Mag* 8:60–88
177. Cheng J, Xu Y, Kong L (2021) Hyperspectral imaging classification based on LBP feature extraction and multimodel ensemble learning. *Comput Electr Eng* 92:107199
178. Zeng X, Yin S-B, Guo Y, Lin J-R, Zhu J-G (2018) A novel semi-supervised feature extraction method and its application in automotive assembly fault diagnosis based on vision sensor data. *Sensors* 18:2545
179. Tejasree G, Agilandeeswari L (2024) A novel multi-class land use/land cover classification using deep kernel attention transformer for hyperspectral images. *Earth Sci Inform* 17:593–616
180. Yu C, Song M, Chang C-I (2018) Band subset selection for hyperspectral image classification. *Remote Sens* 10:113
181. Sun W, Du Q (2019) Hyperspectral band selection: A review. *IEEE Geosci Remote Sens Mag* 7:118–139
182. Cariou C, Chehdi K, Le Moan S (2010) BandClust: An unsupervised band reduction method for hyperspectral remote sensing. *IEEE Geosci Remote Sens Lett* 8:565–569
183. Zhang M, Ma J, Gong M (2017) Unsupervised hyperspectral band selection by fuzzy clustering with particle swarm optimization. *IEEE Geosci Remote Sens Lett* 14:773–777
184. Cao X, Wu B, Tao D, Jiao L (2016) Automatic band selection using spatial-structure information and classifier based clustering. *IEEE J Sel Top Earth Obs Remote Sensing* 9:4352–4360
185. Ahmad M, Haq DIU, Mushtaq Q, Sohaib M (2011) A new statistical approach for band clustering and band selection using K-means clustering. *Int J Eng Technol* 3:606–614
186. Qian Y, Yao F, Jia S (2009) Band selection for hyperspectral imagery using affinity propagation. *IET Comput Vis* 3:213
187. Hedjam R, Cheriet M (2012) Hyperspectral band selection based on graph clustering. In: *2012 11th international conference on information science, signal processing and their applications (ISSPA)*, pp 813–817

188. Archibald R, Fann G (2007) Feature selection and classification of hyperspectral images with support vector machines. *IEEE Geosci Remote Sens Lett* 4:674–677
189. Zare A, Gader P (2008) Hyperspectral band selection and endmember detection using sparsity promoting priors. *IEEE Geosci Remote Sens Lett* 5:256–260
190. Chang CI, Liu KH (2013) Progressive band selection of spectral unmixing for hyperspectral imagery. *IEEE Trans Geosci Remote Sens* 52:2002–2017
191. Chang C-I, Du Q, Sun T-L, Althouse ML (1999) A joint band prioritization and band-decorrelation approach to band selection for hyperspectral image classification. *IEEE Trans Geosci Remote Sens* 37:2631–2641
192. Bajcsy P, Groves P (2004) Methodology for hyperspectral band selection. *Photogramm Eng Remote Sens* 70:793–802
193. Kim J-H, Kim J, Yang Y, Kim S, Kim HS (2017) Covariance-based band selection and its application to near-real time hyperspectral target detection. *Opt Eng* 56:053101
194. Chang C-I, Wang S (2006) Constrained band selection for hyperspectral imagery. *IEEE Trans Geosci Remote Sens* 44:1575–1585
195. Guo B, Gunn SR, Damper RI, Nelson JD (2006) Band selection for hyperspectral image classification using mutual information. *IEEE Geosci Remote Sens Lett* 3:522–526
196. Ifarraguerri A, Prairie MW (2004) Visual method for spectral band selection. *IEEE Geosci Remote Sens Lett* 1:101–106
197. He Y, Liu D, Yi S (2010) Recursive spectral similarity measure-based band selection for anomaly detection in hyperspectral imagery. *J Opt* 13:015401
198. Cao X, Li X, Li Z, Jiao L (2017) Hyperspectral band selection with objective image quality assessment. *Int J Remote Sens* 38:3656–3668
199. Kamandar M, Ghassemian H (2011) Maximum relevance, minimum redundancy band selection for hyperspectral images. In: 2011 19th Iranian conference on electrical engineering, pp 1–5
200. Wang B, Wang X, Chen Z (2012) Spatial entropy based mutual information in hyperspectral band selection for supervised classification. *Int J Numer Anal* 9
201. Patra S, Barman B (2021A) novel dependency definition exploiting boundary samples in rough set theory for hyperspectral band selection. *Appl Soft Comput* 99:106944
202. Tejasree G, Agilandeeswari L (2024) Land use/land cover (LULC) classification using deep-LSTM for hyperspectral images. *Egypt J Remote Sens Space Sci* 27:52–68
203. Du Q, Yang H (2008) Similarity-based unsupervised band selection for hyperspectral image analysis. *IEEE Geosci Remote Sens Lett* 5:564–568
204. Santos LCB, Guimaraes SJF, Santos JAD (2015) Efficient unsupervised band selection through spectral rhythms. *IEEE J Sel Top Signal Process* 9:1016–1025
205. Wang L, Jia X, Zhang Y (2007) A novel geometry-based feature-selection technique for hyperspectral imagery. *IEEE Geosci Remote Sens Lett* 4:171–175
206. Keshava N (2004) Distance metrics and band selection in hyperspectral processing with applications to material identification and spectral libraries. *IEEE Trans Geosci Remote Sens* 42:1552–1565
207. Yang H, Du Q, Su H, Sheng Y (2010) An efficient method for supervised hyperspectral band selection. *IEEE Geosci Remote Sens Lett* 8:138–142
208. Yang H, Du Q, Chen G (2012) Particle swarm optimization-based hyperspectral dimensionality reduction for urban land cover classification. *IEEE J Sel Top Appl Earth Obs Remote Sens* 5:544–554
209. Li S, Zhu Y, Wa D, Feng J (2013) Spectral similarity-preserving hyperspectral band selection. *Remote Sens Lett* 4:969–978
210. Serpico SB, Bruzzone L (2001) A new search algorithm for feature selection in hyperspectral remote sensing images. *IEEE Trans Geosci Remote Sens* 39:1360–1367
211. Backer SD, Kempeneers P, Debruyne W, Scheunders P (2005) A band selection technique for spectral classification. *IEEE Geosci Remote Sens Lett* 2:319–323
212. Huang R, Li X (2008) Band selection based on evolution algorithm and sequential search for hyperspectral classification. In: 2008 international conference on audio, language and image processing, pp 1270–1273
213. Su H, Yong B, Du Q (2015) Hyperspectral band selection using improved firefly algorithm. *IEEE Geosci Remote Sens Lett* 13:68–72
214. Ghosh A, Datta A, Ghosh S (2013) Self-adaptive differential evolution for feature selection in hyperspectral image data. *Appl Soft Comput* 13:1969–1977
215. Gao J, Du Q, Gao L, Sun X, Zhang B (2014) Ant colony optimization-based supervised and unsupervised band selections for hyperspectral urban data classification. *J Appl Remote Sens* 8:085094
216. Zhang L, Zhong Y, Huang B, Gong J, Li P (2007) Dimensionality reduction based on clonal selection for hyperspectral imagery. *IEEE Trans Geosci Remote Sens* 45:4172–4186

217. Nakamura RY, Fonseca LMG, Dos Santos JA, Torres RDS, Yang XS, Papa JP (2013) Nature-inspired framework for hyperspectral band selection. *IEEE Trans Geosci Remote Sens* 52:2126–2137
218. Sawant SS, Prabukumar M, Loganathan A, Alenizi FA, Ingaleswar S (2022) Multi-objective multi-verse optimizer based unsupervised band selection for hyperspectral image classification. *Int J Remote Sens* 43:3990–4024
219. Sun K, Geng X, Ji L (2014) A new sparsity-based band selection method for target detection of hyperspectral image. *IEEE Geosci Remote Sens Lett* 12:329–333
220. Li J-M, Qian Y-T (2011) Clustering-based hyperspectral band selection using sparse nonnegative matrix factorization. *J Zhejiang Univ Sci C* 12:542–549
221. Li S, Qi H (2011) Sparse representation based band selection for hyperspectral images. In: 2011 18th IEEE international conference on image processing, pp 2693–2696
222. Guo Z, Yang H, Bai X, Zhang Z, Zhou J (2013) Semi-supervised hyperspectral band selection via sparse linear regression and hypergraph models. In: 2013 IEEE international geoscience and remote sensing symposium-IGARSS, pp 1474–1477
223. Yin J, Wang Y, Zhao Z (2010) Optimal band selection for hyperspectral image classification based on inter-class separability. In 2010 Symposium on Photonics and Optoelectronics 1–4
224. Stavrakoudis DG, Galidaki GN, Gitas IZ, Theocharis JB (2011) A genetic fuzzy-rule-based classifier for land cover classification from hyperspectral imagery. *IEEE Trans Geosci Remote Sens* 50:130–148
225. Su H, Yao W, Wu Z, Zheng P, Du Q (2021) Kernel low-rank representation with elastic net for China coastal wetland land cover classification using GF-5 hyperspectral imagery. *ISPRS J Photogramm Remote Sens* 171:238–252
226. Guo Y, Yin X, Zhao X, Yang D, Bai Y (2019) Hyperspectral image classification with SVM and guided filter. *EURASIP J Wireless Commun Netw* 2019:1–9
227. Shambulinga M, Sadashivappa G (2019) Hyperspectral image classification using support vector machine with guided image filter. *Int J Adv Comput Sci Appl* 10
228. Lv W, Wang X (2020, 2020) Overview of hyperspectral image classification. *J Sens*
229. Chakravarty S, Paikaray BK, Mishra R, Dash S (2021) Hyperspectral image classification using spectral angle mapper. In: 2021 IEEE international women in engineering (wie) conference on electrical and computer engineering (WIECON-ECE), pp 87–90
230. Yang CC, Prasher SO, Enright P, Madramootoo C, Burgess M, Goel PK, Callum I (2003) Application of decision tree technology for image classification using remote sensing data. *Agric Syst* 76:1101–1117
231. Huang R, Zhu J (2013) Using random forest to integrate Lidar data and hyperspectral imagery for land cover classification. In: 2013 IEEE international geoscience and remote sensing symposium-IGARSS, pp 3978–3981
232. Bittencourt HR, Clarke RT (2004) Feature selection by using classification and regression trees (CART). *The International Archives of the Photogrammetry, Remote Sens Spat Inf Sci*
233. Guo Y, Han S, Li Y, Zhang C, Bai Y (2018) K-Nearest Neighbor combined with guided filter for hyperspectral image classification. *Proc Comput Sci* 129:159–165
234. Addini PF, Hadi W, Harahap PMR (2023) Application of the multivariate adaptive regression spline (Mars) method in analyzing misclassification of elementary school accreditation data in the city of Tebing Tinggi. *Jurnal Scientia* 12:617–620
235. Li W, Prasad S, Tramel EW, Fowler JE, Du Q (2014) Decision fusion for hyperspectral image classification based on minimum-distance classifiers in the wavelet domain. In: 2014 IEEE China Summit & International Conference on Signal and Information Processing (ChinaSIP), pp 162–165
236. Zhao Y, Yuan Y, Wang Q (2019) Fast spectral clustering for unsupervised hyperspectral image classification. *Remote Sens* 11:399
237. Ali UME, Hossain MA, Islam MR (2019) Analysis of PCA based feature extraction methods for classification of hyperspectral image. In: 2019 2nd international conference on innovation in engineering and technology (ICIET), pp 1–6
238. Kaarna A, Toivanen P, Keränen P (2006) Compression and classification methods for hyperspectral images. *Pattern Recognit Image Anal* 16:413–424
239. Zhang L, Zhang L, Du B, You J, Tao D (2019) Hyperspectral image unsupervised classification by robust manifold matrix factorization. *Inf Sci* 485:154–169
240. Patel U, Patel V (2023) A comprehensive review: active learning for hyperspectral image classifications. *Earth Sci Inform* 16:1975–1991
241. El\_Rahman SA (2016) Big data analysis: hyperspectral image processing for agriculture applications. *Int J Comput Digit Syst* 5
242. Bilgin G, Erturk S, Yildirim T (2008) Unsupervised classification of hyperspectral-image data using fuzzy approaches that spatially exploit membership relations. *IEEE Geosci Remote Sens Lett* 5:673–677



243. Paoletti ME, Haut JM, Plaza J, Plaza A, Vigo-Aguar J (2017) Yinyang K-means clustering for hyperspectral image analysis. In: proceedings of the 17th international conference on computational and mathematical methods in science and engineering, pp 1625–1636
244. Gillis N, Kuang D, Park H (2014) Hierarchical clustering of hyperspectral images using rank-two nonnegative matrix factorization. *IEEE Trans Geosci Remote Sens* 53:2066–2078
245. Ren J, Wang R, Liu G, Feng R, Wang Y, Wu W (2020) Partitioned relief-F method for dimensionality reduction of hyperspectral images. *Remote Sens* 12:1104
246. Ramadhani F, Zarlis M, Suwilo S (2020) Improve BIRCH algorithm for big data clustering. *IOP Conf Ser Mater Sci Eng* 725:012090
247. Kuo BC, Huang WC, Liu HC, Tseng SC (2008) A novel fuzzy C-means method for hyperspectral image classification. In: *IGARSS 2008-2008 IEEE international geoscience and remote sensing symposium 2:II-1002*
248. Hung CC, Kulkarni S, Kuo BC (2010) A new weighted fuzzy c-means clustering algorithm for remotely sensed image classification. *IEEE J Sel Top Signal Process* 5:543–553
249. Li X, Ma F, Hu J, Jivkov AP, Chu D (2023) A spatiotemporal identification method for deformation characteristics of expansive soil canal slope based on spectral clustering. *Exp Syst Appl* 225:120108
250. Wahyuningrum T, Khomsah S, Suyanto S, Meliana S, Yunanto PE, Al Maki WF (2021) improving clustering method performance using K-Means, mini batch K-Means, BIRCH and spectral. In: *2021 4th International Seminar on Research of Information Technology and Intelligent Systems (ISRITI)*, pp 206–210
251. Huang X, Zhang L (2008) An adaptive mean-shift analysis approach for object extraction and classification from urban hyperspectral imagery. *IEEE Trans Geosci Remote Sens* 46:4173–4185
252. Bahraini T, Azimpour P, Yazdi HS (2021) Modified-mean-shift-based noisy label detection for hyperspectral image classification. *Comput Geosci* 155:104843
253. Lange J, Cavallaro G, Götz M, Erlingsson E, Riedel M (2018) The influence of sampling methods on pixel-wise hyperspectral image classification with 3D convolutional neural networks. In: *IGARSS 2018-2018 IEEE international geoscience and remote sensing symposium*, pp 2087–2090
254. Datta A, Ghosh S, Ghosh A (2015) Combination of clustering and ranking techniques for unsupervised band selection of hyperspectral images. *IEEE J Sel Top Appl Earth Obs Remote Sens* 8:2814–2823
255. Wu H, Prasad S (2017) Semi-supervised deep learning using pseudo labels for hyperspectral image classification. *IEEE Trans Image Process* 27:1259–1270
256. Maulik U, Chakraborty D (2013) Learning with transductive SVM for semisupervised pixel classification of remote sensing imagery. *ISPRS J Photogramm Remote Sens* 77:66–78
257. Sawant SS, Prabukumar M (2020) A review on graph-based semi-supervised learning methods for hyperspectral image classification. *Egypt J Remote Sens Space Sci* 23:243–248
258. Castillejo-González IL, López-Granados F, García-Ferrer A, Peña-Barragán JM, Jurado-Expósito M, de la Orden MS, González-Audicana M (2009) Object-and pixel-based analysis for mapping crops and their agro-environmental associated measures using QuickBird imagery. *Comput Electron Agric* 68:207–215
259. Xu X, Zhong Y, Zhang L, Zhang H (2012) Sub-pixel mapping based on a MAP model with multiple shifted hyperspectral imagery. *IEEE J Sel Top Appl Earth Obs Remote Sens* 6:580–593
260. Chen N, Zhou H (2020) Classification of hyperspectral image based on superpixel segmentation and DPC algorithm. In: *2020 19th international symposium on distributed computing and applications for business engineering and science (DCABES)*, pp 138–141
261. Yu H, Gao L, Liao W, Zhang B, Pižurica A, Philips W (2017) Multiscale superpixel-level subspace-based support vector machines for hyperspectral image classification. *IEEE Geosci Remote Sens Lett* 14:2142–2146
262. Müller B, Reinhardt J, Strickland MT (2012) *Neural networks: an introduction*. Springer Science & Business Media
263. Anderson JA (1995) *An introduction to neural networks*. MIT press
264. Guo AJ, Zhu F (2018) Spectral-spatial feature extraction and classification by ANN supervised with center loss in hyperspectral imagery. *IEEE Trans Geosci Remote Sens* 57:1755–1767
265. Meng Z, Zhao F, Liang M (2021) SS-MLP: A novel spectral-spatial MLP architecture for hyperspectral image classification. *Remote Sens* 13:4061
266. He X, Chen Y (2021) Modifications of the multi-layer perceptron for hyperspectral image classification. *Remote Sens* 13:3547
267. Li S, Song W, Fang L, Chen Y, Ghamisi P, Benediktsson JA (2019) Deep learning for hyperspectral image classification: An overview. *IEEE Trans Geosci Remote Sens* 57:6690–6709
268. Aberna P, Agilandeswari L (2024) Digital image and video watermarking: methodologies, attacks, applications, and future directions. *Multimed Tools Appl* 83:5531–5591
269. Mou L, Ghamisi P, Zhu XX (2017) Deep recurrent neural networks for hyperspectral image classification. *IEEE Trans Geosci Remote Sens* 55:3639–3655

270. Viel F, Maciel RC, Seman LO, Zeferino CA, Bezerra EA, Leithardt VRQ (2023) Hyperspectral image classification: An analysis employing CNN, LSTM, transformer, and attention mechanism. *IEEE Access* 11:24835–24850
271. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Polosukhin I (2017) Attention is all you need. *Adv Neural Inf Process Syst* 30
272. Xu Y, Du B, Zhang L (2021) Self-attention context network: Addressing the threat of adversarial attacks for hyperspectral image classification. *IEEE Trans Image Process* 30:8671–8685
273. Sun J, Zhang J, Gao X, Wang M, Ou D, Wu X, Zhang D (2022) Fusing spatial attention with spectral-channel attention mechanism for hyperspectral image classification via encoder–decoder networks. *Remote Sens* 14:1968
274. Li Z, Cui X, Wang L, Zhang H, Zhu X, Zhang Y (2021) Spectral and spatial global context attention for hyperspectral image classification. *Remote Sens* 13:771
275. He K, Sun W, Yang G, Meng X, Ren K, Peng J, Du Q (2022) A dual global–local attention network for hyperspectral band selection. *IEEE Trans Geosci Remote Sens* 60:1–13
276. Xue Z, Yu X, Liu B, Tan X, Wei X (2021) HRResNetAM: Hierarchical residual network with attention mechanism for hyperspectral image classification. *IEEE J Sel Top Appl Earth Obs Remote Sens* 14:3566–3580
277. Hong D, Han Z, Yao J, Gao L, Zhang B, Plaza A, Chanussot J (2021) SpectralFormer: Rethinking hyperspectral image classification with transformers. *IEEE Trans Geosci Remote Sens* 60:1–15
278. Yang J, Du B, Wu C (2022) Hybrid vision transformer model for hyperspectral image classification. In: *IGARSS 2022-2022 IEEE international geoscience and remote sensing symposium*, pp 1388–1391
279. Zhang Z, Li T, Tang X, Hu X, Peng Y (2022) CAEVT: Convolutional autoencoder meets lightweight vision transformer for hyperspectral image classification. *Sensors* 22:3902
280. Hu X, Li T, Zhou T, Liu Y, Peng Y (2021) Contrastive learning based on transformer for hyperspectral image classification. *Appl Sci* 11:8670
281. Wang Y, Huang S, Liu D, Wang B (2012) Research advance on band selection-based dimension reduction of hyperspectral remote sensing images. In: *2012 2nd international conference on remote sensing, environment and transportation engineering*, pp 1–4
282. Dong C, Loy CC, He K, Tang X (2014) Learning a deep convolutional network for image super-resolution. In: *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6–12, 2014, Proceedings, Part IV, vol 13*, pp 184–199
283. Roy SK, Krishna G, Dubey SR, Chaudhuri BB (2019) HybridSN: Exploring 3-D–2-D CNN feature hierarchy for hyperspectral image classification. *IEEE Geosci Remote Sens Lett* 17:277–281
284. Ji S, Xu W, Yang M, Yu K (2012) 3D convolutional neural networks for human action recognition. *IEEE Trans Pattern Anal Mach Intell* 35:221–231
285. Navin MS, Agilandeewari L (2020) Comprehensive review on land use/land cover change classification in remote sensing. *J Spectr Imaging* 9:a8
286. MohanRajan SN, Loganathan A, Manoharan P (2020) Survey on land use/land cover (LU/LC) change analysis in remote sensing and GIS environment: Techniques and challenges. *Environ Sci Pollut Res* 27:29900–29926
287. Bangelesa F, Adam E, Knight J, Dhau I, Ramudzuli M, Mokotjomela TM (2020) Predicting soil organic carbon content using hyperspectral remote sensing in a degraded mountain landscape in lesoto. *Appl Environ Soil Sci* 2020:1–11
288. MohanRajan SN, Loganathan A (2023) A novel fuzzy Harris hawks optimization-based supervised vegetation and bare soil prediction system for Javadi Hills, India. *Arab J Geosci* 16:478
289. Navin MS, Agilandeewari L (2020) Multispectral and hyperspectral images based land use/land cover change prediction analysis: an extensive review. *Multimed Tools Appl* 79:29751–29774
290. MohanRajan SN, Loganathan A (2021) Modelling spatial drivers for LU/LC change prediction using hybrid machine learning methods in Javadi Hills, Tamil Nadu, India. *J Indian Soc Remote Sens* 49:913–934
291. Lu Q, Si W, Wei L, Li Z, Xia Z, Ye S, Xia Y (2021) Retrieval of water quality from UAV-borne hyperspectral imagery: A comparative study of machine learning algorithms. *Remote Sens* 13:3928
292. MohanRajan SN, Loganathan A, Manoharan P, Alenizi FA (2024) Fuzzy Swin transformer for land use/land cover change detection using LISS-III Satellite data. *Earth Sci Inform*:1–20
293. Mahmoudi MR, Baroumand S (2022) Modeling the stochastic mechanism of sensor using a hybrid method based on seasonal autoregressive integrated moving average time series and generalized estimating equations. *ISA Trans* 125:300–305
294. Selige T, Böhner J, Schmidhalter U (2006) High resolution topsoil mapping using hyperspectral image and field data in multivariate regression modeling procedures. *Geoderma* 136:235–244
295. Etemadi H, Smoak JM, Karami J (2018) Land use change assessment in coastal mangrove forests of Iran utilizing satellite imagery and CA–Markov algorithms to monitor and predict future change. *Environ Earth Sci* 77:1–13

296. Yang W, Nigon T, Hao Z, Paiao GD, Fernández FG, Mulla D, Yang C (2021) Estimation of corn yield based on hyperspectral imagery and convolutional neural network. *Comput Electron Agric* 184:106092
297. Song YQ, Zhao X, Su HY, Li B, Hu YM, Cui XS (2018) Predicting spatial variations in soil nutrients with hyperspectral remote sensing at regional scale. *Sensors* 18:3086
298. Panda SS, Ames DP, Panigrahi S (2010) Application of vegetation indices for agricultural crop yield prediction using neural network techniques. *Remote Sens* 2:673–696
299. Yang X, Ye Y, Li X, Lau RY, Zhang X, Huang X (2018) Hyperspectral image classification with deep learning models. *IEEE Trans Geosci Remote Sens* 56:5408–5423
300. Zhou X, Sun J, Tian Y, Lu B, Hang Y, Chen Q (2020) Development of deep learning method for lead content prediction of lettuce leaf using hyperspectral images. *Int J Remote Sens* 41:2263–2276
301. Lee H, Kwon H (2016) Contextual deep CNN based hyperspectral classification. In: 2016 IEEE international geoscience and remote sensing symposium (IGARSS), pp 3322–3325
302. He X, Chen Y (2019) Optimized input for CNN-based hyperspectral image classification using spatial transformer network. *IEEE Geosci Remote Sens Lett* 16:1884–1888
303. Tang X, Meng F, Zhang X, Cheung YM, Ma J, Liu F, Jiao L (2020) Hyperspectral image classification based on 3-D octave convolution with spatial–spectral attention network. *IEEE Trans Geosci Remote Sens* 59:2430–2447
304. Chen H, Miao F, Chen Y, Xiong Y, Chen T (2021) A hyperspectral image classification method using multifeature vectors and optimized KELM. *IEEE J Sel Top Appl Earth Obs Remote Sens* 14:2781–2795
305. Chen W, Ouyang S, Yang J, Li X, Zhou G, Wang L (2022) JAGAN: A framework for complex land cover classification using Gaofen-5 AHSI images. *IEEE J Sel Top Appl Earth Obs Remote Sens* 15:1591–1603
306. Deborah H, Richard N, Hardeberg JY (2015) A comprehensive evaluation of spectral distance functions and metrics for hyperspectral image processing. *IEEE J Sel Top Appl Earth Obs Remote Sens* 8:3224–3234
307. Kabisch N, Selsam P, Kirsten T, Lausch A, Bumberger J (2019) A multi-sensor and multi-temporal remote sensing approach to detect land cover change dynamics in heterogeneous urban landscapes. *Ecol Indic* 99:273–282
308. Taufik A, Syed Ahmad SS, Azmi EF (2019) Classification of Landsat 8 satellite data using unsupervised methods. In: *Intelligent and interactive computing: Proceedings of IIC*, vol 2018, pp 275–284
309. Tejasree G, Agilandeewari L (2022) Gradient boosting ensemble method for in-vivo brain tumour classification using hyperspectral images. *Indian J Comp Sci Eng* 13:1600–1672
310. Yang J, Zhao Y, Chan JCW, Yi C (2016) Hyperspectral image classification using two-channel deep convolutional neural network. In: 2016 IEEE international geoscience and remote sensing symposium (IGARSS), pp 5079–5082

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.