



A comprehensive review: active learning for hyperspectral image classifications

Usha Patel¹ · Vibha Patel²

Received: 8 August 2022 / Accepted: 23 June 2023 / Published online: 20 July 2023
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2023

Abstract

Advanced Hyperspectral image sensors can capture high-resolution land cover images. Many supervised Machine learning (ML) and Deep learning (DL) algorithms succeeded in the Hyperspectral image classification for various applications. Scientific findings reveal that supervised learning methods' performance heavily depends on training set size, i.e., labelled by the ground truth collection. Labelling a large volume of hyperspectral images is tedious and time-consuming. Research in Active learning (AL) aims to determine the minimal set of samples for labelling, ensuring that model performance is unaffected by training using a minimal set of samples. This article examined fundamental AL and cutting-edge methods used for hyperspectral image classifications. This article also focuses on the cutting-edge approach known as Active Deep Learning (ADL). ADL combines the strong discriminative capabilities of the deep learning model with active learning. This article examined the use of ADL for the Hyperspectral images (HSIs) classifications and reviewed opportunities and challenges. At last, the experiment illustrates and evaluates the integration of the DL model with several AL approaches.

Keywords Hyperspectral image classification · Machine learning · Active learning · Deep active learning

Introduction

Hyperspectral imaging records spectral reflectance of visible and near-infrared bands in remote sensing. Hyperspectral image sensors provide hundreds of spectral band reflectances of the earth's surface (Khan et al. 2018). These narrow and continuous spectral bands can efficiently discriminate the pixels of land cover images. Many applications like agriculture, geology, forestry, and urban planning use HSIs data for resource management (Adão et al. 2017; Gao et al. 2021; Kahraman and Bacher 2021). With advanced technology, high-resolution hyperspectral imaging sensors can capture

earth observation as digital imagery in HSIs cubes. Effective processing techniques are required to extract valuable information from this enormous volume of data.

Hyperspectral image classification aims to categorize each pixel vector of an image into specific classes corresponding to the image content. Due to its diverse applications, HSI classification has received significant attention over the past years. Figure 1 shows statistics of scientific articles published in the last six years. The statistics collected from the Google Scholar search engine with the keywords: "Hyperspectral image classification." is included in the article's title. According to statistics, the research community has increasingly focused on hyperspectral image classification. HSIs classification techniques presented in the literature broadly cover spectral-based and spectral-spatial-based classification algorithms. The spectral-based approach uses spectral reflectance for the classification, whereas spectral-spatial-based algorithms consider the spatial features along with the spectral reflectance to use the spatial correlation.

Various unsupervised Machine learning algorithms, such as k-means clustering (Zhang et al. 2016a) and fuzzy c-means clustering (Salem et al. 2016), can identify patterns from unlabeled data for classification. Supervised

Communicated by: H. Babaie

✉ Usha Patel
ushapatel@nirmauni.ac.in
Vibha Patel
vibhadp@gmail.com

¹ CSE Department, Institute of Technology, Nirma University, Ahmedabad, India

² IT Department, Vishwakarma Government Engineering College, Gujarat Technological University, Ahmedabad, India

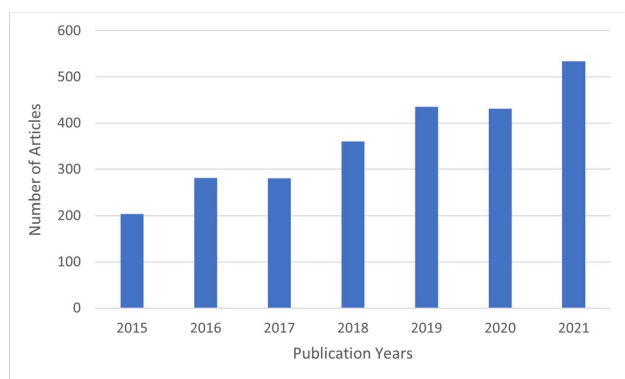


Fig. 1 Statistics of published articles title included "Hyperspectral image classification." according to Google Scholar

ML algorithms like the random forest (Zhang et al. 2018), logistic regression (Qian et al. 2012), and support vector machines (Moughal 2013) that required labelled samples for the training are also widely used and have given promising results for the HSIs classification. Due to the constant evolution of Machine learning, Deep learning, the subdomain of ML, has gained more attention due to its success in various fields. Supervised DL techniques have received a lot of attention for the classification of hyperspectral images. Deep learning can automatically learn hidden patterns from the data, so the DL model received more attention than the other supervised ML models. The main challenges of supervised HSIs classifications include higher dimensionality or redundancy, large spectral and spatial variability and lack of labelled data (Datta et al. 2022). Various dimensionality reduction approaches exist to reduce dimensionality; Principle Component Analysis (PCA) is one of the most popular methods because of its ease of use and effectiveness. Changes in illumination, atmosphere, and environmental conditions can generate spatial and spectral signal variations. Spatial filtering, spatial normalization, and spectral unmixing are the techniques for dealing with spectral and spatial variability. The performance of the supervised machine learning model depends on the quantity and the quality of the labelled dataset used for the model training (Yang et al. 2018). Specifically, the DL models need more labelled samples to train great number of learning parameters. This article discusses current strategies put forth by the research communities to address the shortage of labelled data for the HSIs classification.

Hyperspectral image classification with limited labelled samples is more attractive because labelled datasets in remote sensing are expensive and time-consuming to produce. The large spatial area and accessibility to locations can make it difficult and costly to acquire suitable training sets through field surveys. As a result, training sets often include fewer samples for the extensive coverage of the scene

(Crawford et al. 2013). Using visual interpretation, one can also acquire labelled data from high-resolution images. But this method is erroneous, subjective, and dependent on the analyst's knowledge. Searching for small data sets with high training utility is worthwhile to lower the cost of human annotation. The performance of the classifier is not affected in this way. Given the availability of unlabeled data, it is also advantageous to exploit the spectrum information in both the labelled and unlabeled data to provide better classification results.

Semisupervised learning, Transfer learning, Few-shot learning, and Active learning are the major categories to overcome the limited labelled training data problem.

Semisupervised learning uses labelled and unlabeled datasets to train models (Sawant and Prabukumar 2020). The three main Semisupervised learning method groups are pseudo labelling (Wu and Prasad 2017; Patel et al. 2020), graph-based methods (Ma et al. 2016), and generative methods (He et al. 2017). Few-shot learning aims to recognize new categories of input samples from a few annotated samples. Few-shot learning significantly improves performance once the deep learning model is tuned (Liu et al. 2018b). Cross-domain (Li et al. 2021) and fusion-based (Liang et al. 2021) few-shot learning was proposed for HSIs classification. Transfer learning, which includes transferring knowledge between two or more related domains, addresses the insufficiency of labelled samples. Literature covers heterogeneous transfer learning (He et al. 2019) and ensemble-based transfer learning (Liu et al. 2020) with DL models for HSIs classifications.

Active learning methods seek to construct representative training data sets for the supervised learning algorithm at the lowest possible cost. It can efficiently acquire discriminative features, with a minimum set of labelled samples, by selecting representative or highly-informative samples for manual labelling from the unlabeled pool. Many primary and advanced AL techniques are applied to mitigate the problem of a limited labelled dataset of HSIs classification. Active deep learning (ADL) is an emerging field that combines deep learning and active learning. This paper focuses on recent articles on active learning, and ADL applied for hyperspectral image classifications.

Contributions

The contributions of this article are as follows:

Survey on active learning: This review summarizes the various active learning techniques used in HSI classifications to overcome the problem of limited labelled samples. The article covers both basic AL techniques and advanced options for improvement.

Active deep learning: Active deep learning is an emerging field capable of extracting discriminative features by the

DL model with limited labelled samples. The comprehensive and insightful surveys of state-of-the-art methods of ADL with HSIs classification are described and thoroughly surveyed.

Comparative analysis of AL methods: This article also emphasizes how the ADL is used to classify HSIs. It also examines the effects of various AL techniques using the same DL frameworks. The impact of the AL parameters is analyzed using the classifier test accuracy results.

Literature reveals comprehensive and insightful surveys on active learning methods; some are specific to remote sensing. For example, the authors in (Thoreau et al. 2022) compared different AL techniques with HSIs classifications. The author implemented various fundamental AL techniques with the same framework (SVM-ML model) to analyze the performance. Jia et al. (Jia et al. 2021) comprehensively review the state-of-the-art deep learning-based methods for HSI classification with few labelled samples. The first article emphasizes the different AL methods, while the second emphasizes the different classifiers to train with limited labelled samples. This article presents a recent survey on active learning and active deep learning for the classification of hyperspectral images.

Organization

Sect. "Active learning" describes active learning and the different categories of AL. Sect. "Advanced Active Learning Techniques" describes advanced AL approaches applied to the HSIs classifications. Sect. "Active deep learning" defines Active deep learning for HSIs classifications and its major problems and solutions. Sect. "Experiment" of the article presents the experimental setup and results of multiple active learning (AL) methods with a deep learning (DL) model as the classifier. Sect. "Conclusion and discussion" is the conclusion and discussion.

Active learning

Active learning (AL) is a method that selects informative samples from an unlabeled dataset, utilizing human input to label these selected samples while training a supervised machine learning model. This approach aims to reduce the reliance on a large labelled training dataset. There are two main types of AL: stream-based and pool-based. In stream-based AL, the algorithm receives each unlabeled sample one at a time and decides whether to query for the label. On the other hand, Pool-based AL involves a sizable pool of unlabeled samples that are presented to an AL acquisition function for selection and manual labelling. In an AL framework, both a supervised machine learning algorithm and an acquisition function play crucial roles (Yang et al.

2018). The acquisition function serves as a query technique to choose informative samples for manual labelling. Figure 2 illustrates the iterative process of pool-based AL, where a limited labelled training dataset trains the supervised machine learning model. It subsequently selects samples from the pool of unlabeled data for annotation based on the model's predictions. This article focuses on discussing the pool-based AL technique.

Active learning techniques

Various AL methods have been applied in different domains in the last two decades. This section will overview the most often used methods and broadly divide them into four categories: posterior probability-based, margin-based, committee-based, and learning-based, as shown in Fig. 3.

- **Posterior probability-based:** Supervised ML models trained by labelled training datasets define the class prediction probability of the unlabeled pool dataset. The predicted posterior probability of unlabeled samples represents the prediction's uncertainty. Uncertain sampling (Lewis and Catlett 1994) is the method to select the most uncertain data points for labelling based on the calculated entropy from the prediction output of the classifier. Mutual information (MacKay 1992) defines a Bayesian framework to measure the effectiveness of the candidate data point. Breaking tie (BT), (Luo et al. 2005) select the candidate sample with a minimum difference of the highest two-class prediction probabilities. Multiclass classification problems are more suitable for this method. The least confidence (LC) (Culotta and McCallum 2005) selects the candidate sample with the least prediction probability.

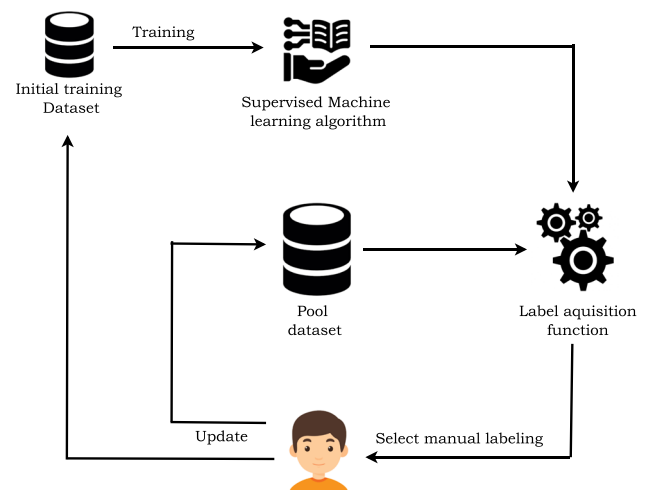
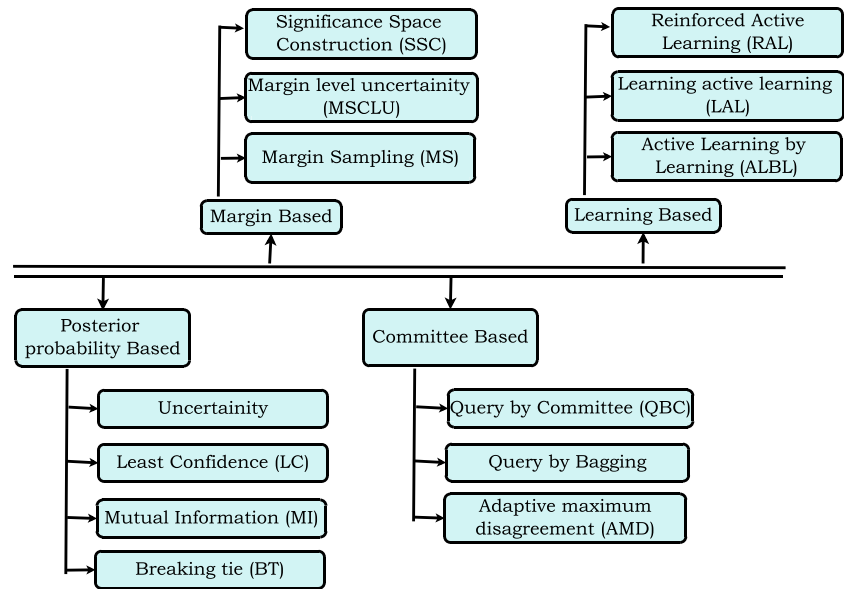


Fig. 2 Pool-based AL framework

Fig. 3 Basic AL methods



- **Margin-based:** Margin-based sampling methods are more suitable with classifiers like Support vector machines (SVM). Margin sampling (Tuia et al. 2011) selects samples for labelling with a minimum distance from the hyperplane. Multiclass level uncertainty (Melgani and Bruzzone 2004; (Demir et al. 2010) is another margin-based sampling method that extends MS for the multiclass classification problem. In this method, the one-against-one and one-against-all strategies are used with multiple SVM classifiers to find the uncertain and diverse samples for the annotation. Significance Space Construction (Pasolli et al. 2010) is a method that constructs significance space to direct the sample selection with the SVM classifier. Best-versus-Second-Best (BvSB) (Joshi et al. 2009) generalizes margin-based multiclass uncertainty compatible with the large data size and multiple class classification.
- **Committee-based:** Committee-based AL method makes a committee and takes decisions based on multiple supervised learning models. Query by Committee (QBC) (Seung et al. 1992) is the first committee-based method to decide the sample selection based on maximum disagreement. The QBC approach is improved by running the algorithm several times, resampling the data, and choosing and picking the point by majority voting (Abe 1998; Tuia et al. 2009). Adaptive maximum disagreement (Leskes 2005; Di and Crawford 2010) is another improved committee-based algorithm considering learner diversity in voting.
- **Learning-based:** Recently, learning-based AL have been proposed to overcome the limitations of heuristics-based AL methods. The performance of the various heuristics-based AL methods depends on the dataset complexity

and the learning model. Due to the success of reinforcement learning (RL) by Atari (Mnih et al. 2013) game playing, many areas use RL to solve the complex problem of different domains. Recently the learning-based AL methods were proposed with the RL framework. Active learning by learning (Hsu and Lin 2015) is an adaptive method that selects the AL method from the different heuristic-based AL techniques based on the K-arm bandit theory. Learning active learning (Konyushkova et al. 2017), a data-driven AL method, can outperform the various heuristics-based AL methods. Reinforced Active Learning (Haußmann et al. 2019) is a policy network for the AL based on deep reinforcement learning.

- **Others:** Bayesian Active Learning by Disagreement (BALD) and Coreset are well-known AL algorithms. BALD (Houlsby et al. 2011) method gains information from predictive entropies, which can work with complex classification models. Many language processing and image processing tasks use deep neural networks. The Coreset (Sener and Savarese 2017) AL method is defined as selecting a core set, which refers to a subset of data points chosen to enable a model trained on this subset to achieve competitive performance on the remaining data points. The Coreset algorithm is designed with a CNN classifier.

Uncertainty-based AL methods decisions rely on the model predictions. Model training with AL is not reliable in its initial phase. So sometimes, uncertainty base AL methods perform worse than random sampling. Margin-based AL method considers the inherent structure of margin-based classifiers like SVM. So these methods are limited by such a classifier. Committee-based methods take decisions with

multiple learners, views, or features. To generate various views or train multiple learners is a computationally intensive task. Due to their adaptability, learning-based AL approaches have recently become more prevalent. However, it's essential to note that these approaches have predominantly been restricted to specific fields, such as language processing, computer vision, and object identification. As of now, classifications of hyperspectral images do not use learning-based AL.

Advanced active learning techniques

This section overviews various advanced active learning (AL) methods for classifying hyperspectral images (HSI). MultiView Active Learning (MVAL), Superpixel-based AL, Cluster-based AL, and Feature-based AL are the broad categories of these methods. Each of these categories offers unique approaches to enhance the efficiency and effectiveness of AL for HSI classification tasks.

MultiView active learning (MVAL)

MultiView Active Learning (MVAL) is a technique that enhances the selection of candidate samples for improved classification performance by leveraging information from multiple views of an input image. By extracting and utilizing information from different perspectives or modalities within the image, MVAL aims to maximize sample diversity and informativeness. This approach ultimately enhances the learning process and improves the accuracy of classification tasks. The set of spectral bands, combinations of spectral-spatial information, texture information, various classifiers output, multiple sensor inputs, or other feature extraction techniques provides different views of the hyperspectral image.

Providing various views of the input image to deliver diverse and complementary information is a critical challenge for MVAL (Crawford et al. 2013). The effectiveness of the MVAL algorithm relies on two assumptions. First, the generated views must be complementary, i.e., there should not be a correlation between the different views. Second, the generated views must be sufficient for any sample in an image to be accurately labelled by one of the views (Muslea et al. 2006).

With the help of clustering, uniform band slicing, and random selection combined with dynamic view updating and feature bagging, Di and Crawford (2011) produced a multiview of the HSIs for classification. Zhou et al. (2016) proposed 3D -redundant wavelet transformation to generate diverse multiple views of an image. To reduce the annotation cost, the author applied intersection-based query selection. The authors also suggested a singularity-based criterion to

include spatial information for better feature extraction of HSIs. Coarseness, content, contrast, smooth component pair, and direction are used as morphological components by Xu et al. (2017) to produce various perspectives of one image. The query strategy uses the cluster distribution of unlabeled samples and uncertainty in classifier prediction.

Hu et al. (2018) defined MVAL for HSIs classification with a 3D Gabor filter to create multiple views of the image. They defined the AL strategy with "internal uncertainty", represented by a classifier's posterior probability, and "external uncertainty", represented by inconsistencies between views. Jamshidpour et al. (2020) proposed multi-view, multi-learner for HSIs classification. The author used a genetic algorithm to create multiple views and different kernels as the multiple learners. With multiple learners, the computational complexity also increases. Xu et al. (2021) proposed a framework to mitigate the inaccurate prediction of classifiers with limited training samples. The designed algorithms "leave-one-class-out" use MVAL with spectral-spatial features for HSIs classification. Candidate samples for the annotation are selected based on the contribution of the training and the level of classification confidence. Cai et al. (2021) proposed phase-induced Gabor filters to generate multiple views of the HSIs for classification. The proposed algorithm can adjust the Gabor filter's frequency response characteristics through the different phase values.

Multiview active learning techniques perform well in the field of HSIs classification. HSIs carries hundreds of spectral band along with high spatial resolutions. Processing large-scale 3D image cubes with multiple views raises computational requirements. Algorithms also need multiple classifiers/learners, adding to the computational complexity.

Superpixel based AL

The superpixel aggregates similar pixels to produce meaningful entities and reduce further processing steps. The literature defines graph-based, watershed-based, density-based, clustering-based, and wavelet-based superpixel algorithms. Stutz et al. (2018) comprehensively reviewed all such superpixel algorithms. Superpixel is widely used in image segmentation to over-segment the image for downstream processing. Superpixels, collections of related pixels for a hyperspectral image, allow for compelling spatial-spectral features for HSIs classification. Priya et al. (2015) illustrates using superpixels and information fusion to extract vital spatial data for classifying HSIs. Guo et al. (2016) suggested Superpixel-based active learning for classifying HSIs. The proposed AL method extracts spectral and spatial features like texture from the superpixel to enhance the AL performance.

Semisupervised active learning is a framework that generates pseudo labels from the actively selected samples.

Liu et al. (2018c) proposed superpixel-based Semisupervised active learning for HSIs classification. The author defined superpixels based on local similarity to generate pseudo-labels for semi-supervised learning. The pseudo-label candidate sample is selected based on the breaking tie AL technique. Lu and Wei (2021) proposed multiscale superpixel-based AL for HSIs classification. The author chooses samples from unlabeled datasets based on diversity and uncertainty in the proposed work. Superpixels are chosen and given pseudo labels among these selected samples, while human experts label the remaining samples. Xue et al. (2018) introduced batch mode AL to choose uncertain and diversified samples based on superpixels for HSIs classification. The candidate selects only one sample per superpixel for the batch to ensure the selection of more diverse samples for the annotation.

Determining the number of superpixels and the superpixel map is a challenging task. The superpixel map highly influences the classification result.

Cluster based AL

Unsupervised machine learning techniques such as clustering allow for identifying groups of related data. The clustering algorithm is attractive since it is quick, easy, and unsupervised. As a result, numerous researchers used a clustering approach to address classification issues with hyperspectral images (Qin et al. 2019; Zhang et al. 2016b), and (Hajiani et al. 2021). Clustering methods can quickly search and identify informative and diverse samples. Zhao et al. (2019) defined a semisupervised clustering-based Generative adversarial net for HSIs classification. Lu et al. (2017) defined the AL framework using committee-based criteria along with BT and spectral clustering to choose an informative and diverse sample selection to train the DL model for the HSIs classification. Patel et al. (2021) selected a sample for annotation using clustering and BvSB active learning technique to train the CNN model with the limited training dataset. Ding et al. (2022) use K-means clustering to define the pseudo label to train a deep neural network. To create a pixel cluster, Dong et al. (2021) retrieved the spectral-spatial characteristics of a pixel. Based on the pixel cluster pseudo label is generated to train the CNN model.

Feature-based AL

Hyperspectral image contains spectral-spatial data cubes. Numerous studies have described a cutting-edge technique for identifying and merging spectral-spatial information, which can significantly boost the performance of AL approaches. The boundary pixels of HSIs are the overlapping regions of different classes where the traditional AL will perform inadequately. Liu et al. (2017) proposed feature-driven

AL where the discriminative features rearrange the original data to target the overlapping region problem. The authors selected features created by the Gabor filter using two crucial feature selection criteria: overall error probability and Fisher ratio. Spatial coordinates are also an essential parameter for pixel classification. Combining spectral and spatial features with spatial coordinates, Mu et al. (2020) extracted the features for HSIs classification. Uncertainty-based AL with SVM classifier selects the samples for manual labelling.

To address the problem of training DNN with a limited training dataset, Deng et al. (2018) propose batch mode AL with deep spectral-spatial features fusions. The author defines separate subnetworks to extract deep spectral and spatial features and fuse them seamlessly. Uncertainty-based AL expands the training set to fine-tune the DNN model. Most heuristics-based AL techniques constrained by the lack of training examples are formed from the classifier's output. To overcome this problem, Wang and Ren (2020) proposed a Generative adversarial network (GAN) to acquire heuristics. The GAN model extracts the high-level features from the intermediate layer of DNN.

Others

This section addresses alternative cutting-edge fuzzy-based, fusion-based, segmentation-based, and adversarial network approaches to improve conventional AL techniques.

An analysis technique, fuzzy logic, was created to include uncertainty in a decision model. Fundamentally, fuzzy logic allows for consideration of approximative rather than precise reasoning. Azadegan et al. (2011) discuss the different applications of fuzzy logic. Ahmad et al. (2020) proposed fuzzy model prediction uncertainty to select distinguished samples for the labelling. The author described fuzziness learning AL as preserving data stability and minimizing selection bias in the spatial domain. Ahmad (2020) proposed Fuzzy-based spectral-spatial discriminant information between and within local and global classes for HSI classification. The author first introduced spatial fuzzy-based misclassification sample information to choose valuable samples.

Utilizing images captured by multiple sensors enhance the accuracy of land cover classifications. Kalita et al. (2021) proposed a cross-sensor adaptation strategy by aerial and hyperspectral image datasets for the land cover classifications. The authors proposed feature extraction with sample stacking that can balance the cross-sensor data and a limited number of labelled samples.

Spatial information carries important information when extracting features of HSIs. Zhang et al. (2015) proposed HSIs classification with hierarchical segmentation as an iterative AL process.

Pixels near the edge of an object are difficult to identify. Samat et al. (2019) proposed the AL technique for resolving

this issue, taking into account edge gradient information along with uncertainty and diversity to choose the most informative sample.

Generative adversarial network (GAN) is becoming more popular and used in many image processing applications. Wang et al. (2022) modify the adversarial autoencoder to extract deep features. Distance between an actual and learned distribution integrated with multi-variance posterior probability is used to identify the candidate sample.

Table 1 summarizes the referred advanced active learning approaches.

Active deep learning

The preceding section describes numerous traditional and cutting-edge methods for classifying HSIs based on AL. This section focuses on the emerging field of Active

Deep Learning, a combination of Deep neural networks and active learning. It is also called the *deepAL* (Deep active learning) (Ren et al. 2021). Usually, the supervised ML model is used in traditional AL approaches. However, DL, a subset of ML, has recently outperformed other ML techniques, particularly for image classification. The ADL studies active learning with the DL model as the learner. Deep learning models, a subset of ML, have been increasingly utilized for hyperspectral image classification, leading to notable advancements in the field. A Deep neural network (DNN) can automatically produce a hierarchical feature representation that is more reliable for classifying HSIs (Zhong et al. 2018). With multilayer nonlinear transformation, the DNN structure may extract meaningful and discriminative features (Feng et al. 2017) for the classification. Many DNN models (Deng et al. 2014; Chen et al. 2016) show significant success in hyperspectral image classification. Convolutional neural network (CNN),

Table 1 Summary of advanced active learning techniques

Ref	Technique	Research insights and future scope
(Zhou et al. 2016)	MVAL	3D -redundant wavelet transform to generate multiple views, intersection-based query selection, the singularity-based criterion for the spatial information
(Xu et al. 2017)	MVAL	Morphological components for multiple views, Optimal feature selection can improve the performance
(Hu et al. 2018)	MVAL	Multiple cubes generated with 3D—Gabor filter for multiple views, posterior probability, and inconsistencies between views as the AL strategy
(Jamshidpour et al. 2020)	MVAL	Genetic algorithm for multiple views and multiple learners for AL
(Xu et al. 2021)	MVAL	Selection of the candidate samples based on training contribution and classification confidence
(Cai et al. 2021)	MVAL	Phase-induced Gabor filters can adjust the frequency response characteristics through P
(Guo et al. 2016)	Superpixel—AL	Extracts the spectral and spatial features like texture from the superpixel, generates pseudo labels from the active selection samples
(Liu et al. 2018c)	Superpixel—AL	Generate pseudo-labels for superpixels based on local similarity
(Lu and Wei 2021)	Superpixel—AL	Samples are selected based on uncertainty and diversity, Superpixels are identified, and pseudo labels generated
(Xue et al. 2018)	Superpixel—AL	Batch mode AL to choose uncertain and diversified samples based on superpixel and neighbourhood information
(Lu et al. 2017)	Clustering	The committee-based AL, BT, and spectral-based clustering for AL
(Patel et al. 2021)	Clustering	Clustering is used for the diversity, BvSB for uncertainty for the sample selection
(Ding et al. 2022)	Clustering	Train a deep neural network with selected samples based on K-means clustering
(Dong et al. 2021)	Clustering	Create a pixel cluster based on spectral-spatial characteristics of a pixel
(Liu et al. 2017)	Feature-based AL	Features created by the Gabor filter are selected by overall error probability and Fisher ratio
(Mu et al. 2020)	Feature-based AL	Spatial coordinates along with spectral and spatial information for feature extraction
(Deng et al. 2018)	Feature-based AL	Separate subnetworks extract deep spectral and spatial features and fuse them seamlessly
(Wang and Ren 2020)	Feature-based AL	Generative adversarial network to acquire AL heuristics
(Ahmad et al. 2020)	Other (fuzzy)	Fuzzy model prediction uncertainty to select distinguished samples for the labelling
(Ahmad 2020)	Other (fuzzy)	Fuzzy-based spectral-spatial information between and within local and global classes
(Kalita et al. 2021)	Other (Sensor fusion)	Aerial and hyperspectral sensor images to extract features
(Zhang et al. 2015)	Other (Segmentation)	Hierarchical segmentation to extract spatial features for iterative AL
(Samat et al. 2019)	Other (Edge gradient)	Edge gradient with AL to identify the pixel near the edge
(Wang et al. 2022)	Other (GAN)	Deep features extracted by adversarial autoencoder for AL-based HSIs classification

Recurrent neural network, and Autoencoder are the DNN models often used for HSIs classifications.

A CNN may extract distinctive features by exploiting spectral and spatial information from the hyperspectral images. 2D convolutional layers (Lee and Kwon 2017; Zhang et al. 2017) and 3D convolutional layers (Fang et al. 2020) are used to extract the spectral and spatial features of HSIs. A recurrent neural network is a DL model with a memory unit that stores sequential information. Spectral-based Long short-term memory (LSTM) (Mou et al. 2017) and spatial-spectral based LSTM (Liu et al. 2018a) are used for the HSIs classifications. Autoencoder is a simple DNN model which consists of an encoder and decoder. The encoder maps the input to the hidden representation, and the decoder maps the hidden representation to the reconstructed output. The autoencoder learning process minimizes the difference between input and reconstructed output. Abdi et al. (2017) and Xing et al. (2016) defined stacked autoencoder to learn distinguished features for the hyperspectral image classification.

ADL: challenges and solutions

Active deep learning is widely employed in a variety of domains. In the literature, the various problems are listed, along with their solutions for ADL. The systematic study discovered the following significant problems, and their solutions as follows:

Challenges:

- **Model uncertainty:** Most AL methods are based on the learner's uncertainty (DL model). The prediction output is initially highly unreliable when the model is trained with very few labelled samples. Due to that, the AL algorithm's performance based on the model prediction sometimes performs worse than the random sampling.
- **Overfitting due to insufficient labelled data:** The classic AL method updates the training set by choosing samples for annotation one at a time. DL models are greedy for the training data. Such little updates to training data could cause the model to become overfit.
- **Processing sequence:** Many AL algorithms are based on feature representation. DL model is a collective process during training, feature extraction, and classification optimization.

In the literature, certain solutions are presented to mitigate the above challenges, which are defined below:

Solutions:

- **Batch mode AL:** Batch mode is the most frequently used solution for ADL in which the learner will query the batch (group) of unlabeled data for the annotation. Add-

ing multiple samples in every iteration will reduce the overfitting of the DL model during training. The batch mode AL will accelerate the learning process of the DL model.

- **Adversarial Network:** The iterative training process in AL is accelerated by the actively selected informative samples. Generative adversarial active learning produces informative synthetic samples instead of querying from the unlabeled pool (Zhu and Bento 2017; Tran et al. 2019). Along with the AL method selected samples, the generated synthetic samples used to train the DL model can reduce the model uncertainty.
- **Active semi-supervised learning:** With a pool of labelled and unlabeled datasets, active semi-supervised learning trains the model. Based on the label of the actively selected informative sample, a pseudo label is generated for the most confident samples from the unlabeled data pool (Rottmann et al. 2018). Along with the labelled training dataset, the pseudo-label generated samples are added to increase the training dataset size to reduce the overfitting problem.

ADL for HSIs classification

In recent literature, the classified ability of different deep learning models has been combined with active learning (AL) techniques to alleviate the requirement for many labelled data. The following section provides a summary of selected scientific literature on this topic.

Lei et al. (2021) defined uncertainty and diversity by the k -means cluster as the candidate sample selection criteria. The author designed a light selector network (DL model) for the uncertainty prediction of the unlabeled dataset and customized loss function. Cross-sensor remote sensing images contain inherent variations in spectral responses and ground sampling distance. Kalita et al. (2021) proposed DL-based cross-sensor domain adaptation for land cover classification with AL. Cross-sensor remote sensing can improve the classification, but at the same time, the complexity of the classifier and the number of labelled sample requirements are also increasing. In the study conducted by Lei et al. (2021) and Kalita et al. (2021), the authors employed a method where uncertain samples were selected, and the classification task was performed separately to stabilize the process. However, it was noted that this approach led to an increase in computational costs. Adversarial networks have also given promising performance in hyperspectral image classifications (Zhu et al. 2018; Wang et al. 2021a). An adversarial autoencoder with a customized AL technique based on dictionary learning and distributional distance was proposed by Wang et al. (2022). The author proposed a query method that used multi-variance in the posterior probability for the uncertainty calculations and the distance between the learned distribution

and the true data distribution for the diversity estimation. The informative and representative measurement used for the candidate sample selection can stabilize and increase accuracy.

Cao et al. (2020a) proposed a 3D discrete wavelet transform to alleviate the different types of noise during HSI classifications. The author used the BvSB active learning method and the CNN classifier. The author used the BvSB AL technique, but the other AL methods can also improve the classification with the limited labelled samples. Wang et al. (2021b) defined customized AL with 3D wavelet features for HSIs change detection. When the multi-temporal HSIs data is acquired with the sensor, certain types of noise can also be introduced, impairing the change detection performance. To overcome this problem author used 3D—wavelet transform features as the input of the CNN model. The customized AL method is also designed to mitigate the problem of the large annotated dataset. Cao et al. (2020b) defined the CNN classifier with AL for HSIs classifications. The Markov random field (MRF) function is also used as the post-processing to enforce the class label smoothness. Data augmentation technique is used to tackle the overfitting problem. The author produced detailed results for better comparisons. Murphy (2020) proposed a diffusion geometry identifier to recognize the homogeneous region of the image to select a sample for the query. Sometimes CNN drops in performance when spatial transformations are applied to the data. CapsNet (Sabour et al. 2017) offers a more detailed representation, reduced spatial variance, and improved interpretability compared to CNNs. Paoletti et al. (2020) used AL-based CapsNet for HSI classifications. The AL iterative process is executed for 80 iterations, and during every iteration, ten samples are selected and, after annotation, added to the training dataset. Deng et al. (2018) proposed active transfer learning by deep joint spectral-spatial feature representation using a stacked sparse autoencoder network to uncover the underlying structure and context of HSIs efficiently. The utilization of more advanced AL techniques can enhance the outcome of classification.

Ahmad et al. (2020) designed a generalized fuzziness extreme learning machine autoencoder with different AL techniques. The proposed algorithm is tested on the Salinas HSIs dataset. The experiment should be run on numerous benchmark datasets to demonstrate the viability of the suggested approach. Lin et al. (2018) takes the HSIs from different sensors and dimensions to extract high-level features for deep active transfer learning. The designed framework is for binary classification, which can also be extended for multiclass classification. Haut et al. (2018) defined different AL techniques with Bayesian CNN classifier with spectral-spatial features extraction. The exhaustive analysis of the different AL techniques with the Bayesian CNN strengthens the article. The computational complexity is significantly

high but can be mitigated by exploring alternative solutions. Liu et al. (2016) proposed a novel AL framework, "weighted incremental dictionary learning," which selects samples based on extracted features from unsupervised RBM and supervised DBN model. The candidate sample selection is based on sparsity estimated with unsupervised learning and uncertainty computed with the supervised learning algorithm. The sparse coding technique has a relatively high computational complexity, resulting in the fusion AL method requiring extensive computation. Pixel-based HSI classification using a Restricted Boltzmann Machine (RBM) classifier and several AL strategies was proposed by Sun et al. (2016). In this experiment, the author solely utilized spectral features as input to the RBM, which could be further enhanced by incorporating spatial information. Samat et al. (2016) defined AL heuristics based on pure and mixed pixels; pure pixels are selected based on the pixel purity index, and the mixed pixel is determined based on distance with the pure candidate. Li (2015) suggested batch mode AL with stacked autoencoders for HSI classification. The author utilized the uncertainty of unlabeled samples as the AL criterion. The difference between the highest two-class prediction probability is used to determine uncertainty. The exhaustive experiment with different datasets and the more result analysis could give better insights. Tuia et al. (2012) proposed hyperspectral image segmentation with queried samples for annotation using unsupervised clustering and active learning with a prune-and-label strategy.

In many methods, batch mode active learning is utilized, which involves selecting a batch of samples based on the same criteria. However, this approach may lead to including similar samples within the batch, thereby limiting the performance of the AL algorithm. Furthermore, employing adaptive query selection criteria instead of a static one during AL iterations has the potential to enhance the output of the active learning algorithm.

Table 2 summarises the referred literature with the details of AL techniques, features used, and the DL model.

Experiment

This section mainly conducts a comprehensive set of experiments from three aspects. First, experiments are designed to demonstrate the advantages of Active deep learning on HSIs classification. Second is the comparative analysis of the classification performance of several frequently used AL methods with a customized CNN model. Third, visualize the impact of batch size in a batch mode ADL for HSIs classifications. Three benchmark HSIs datasets are used to complete our experiments, i.e., Indian Pines, Pavia University, and Salinas Valley.

Table 2 Summary of Active Deep Learning for HSIs classification

Ref	AL method	Specialized features	Classifier	Research insights
(Lei et al. 2021)	Uncertainty and diversity by K-means	Spectral, spatial features from 3D patch	ResNet Architecture	Defined selector network for predicting uncertainty of unlabelled samples with a customized loss function
(Kalita et al. 2021)	Standard deviation-based AL	Features from source and target domain	Deep CNN	Cross-sensor remote sensing with domain adaptation-based DL model for land cover classification
(Wang et al. 2022)	Customized AL	-	Adversarial autoencoder	Sample selection by the multi-variance in the posterior probability and the distance between the learned and true distribution
(Cao et al. 2020a)	Best-versus-second best	Enhanced 3DDWT feature	CNN	Enhanced 3D discrete wavelet transform to mitigate the stripe noise and extract spectral-spatial features
(Wang et al. 2021b)	Pixel with the largest probability of changing class	3D-Wavelet transform	CNN	With active learning, change detection from the HSIs with CNN and 3D- wavelet features
(Cao et al. 2020b)	Random sampling, Entropy sampling, and BvSB	Spectral-spatial features from the 3D patch	CNN	CNN model with AL for HSIs classifications
(Paoletti et al. 2020)	Random sampling, Breaking-tie	Spectral and the spatial-contextual information	CapsNets	AL-based CapsNet for HSI classifications to overcome the limitations of CNN
(Deng et al. 2018)	Multiclass-level uncertainty, RS, MS	Stacked spectral-spatial features	Stacked sparse autoencoder	Active transfer learning with the stacked sparse autoencoder network to uncover the underlying structure and context of HSIs
(Ahmad et al. 2020)	RS, BT, MBT, Mutual information	Spectral-spatial from 3D patch	Autoencoders	Active learning with Generalized fuzziness extreme learning machine autoencoder defined for HSI classification
(Lin et al. 2018)	Hybrid (Informative and Representative)	High-level features from source and target domains	Autoencoders	HSIs with different sensors and dimensions, deep active transfer learning with extracted high-level feature correlation
(Haut et al. 2018)	RS, Mean STD, ES, BALD, BT, MS	spectral and spatial $\hat{\alpha}$ "contextual information	Bayesian CNN	AL-based Bayesian CNN classifier to extract spectral-spatial features of HSIs
(Liu et al. 2016)	Weighted incremental dictionary learning	Unsupervised and supervised DNN	Deep belief network	An AL algorithm, "weighted incremental dictionary learning," is based on the extracted features from unsupervised RBM and supervised DBN model
(Sun et al. 2016)	RS, Mutual information, BT	Pixel-wise extracted features	RBM	Active learning with autoencoder for HSI classification
(Li 2015)	Uncertainty of NN	Redundant spatial and spectral features	Stacked autoencoder	Batch mode AL, the difference between the highest two-class prediction probability to calculate the uncertainty

Table 3 Dataset Summary

Dataset	Sensor	Samples	Classes	Spectral Band	Spatial resolution
Indian Pines	AVIRIS	10,249	16	200	20 Meter
Pavia University	ROSIS	42,776	9	103	1.3 Meter
Salinas Valley	AVIRIS	54,129	16	204	3.7 Meter

Dataset

Indian Pines (IP), Pavia University (PV), and Salinas Valley (SL) are the publicly available hyperspectral images captured by airborne sensors. The IP dataset image, 145 Å— 145 pixels and 224 spectral reflectance bands was obtained with an AVIRIS sensor across a region of northwest Indiana. After removing water absorption bands, a 200 spectral band is used for further processing. Every pixel of an image represents 20 m of the earth's surface as the spatial resolution. In this image, a total of 10,249 pixels contains the corresponding ground truth, i.e. the labelled samples. A total of 16 classes are represented in this image, each with a different type of crop. The PV image is captured with the ROSIS sensor with 610 X 610 pixels over Pavia University, Italy. This image contains 102 spectral band reflectance with 1.3 m spatial resolutions. This image includes nine classes with 42,776 labelled samples. The AVIRIS sensor over Salinas Valley, California, takes the SL image. This image has 3.7 m spatial resolution with 204 spectral reflectance bands after removing noisy bands. This dataset has 54,129 labelled samples of 16 different classes. Table 3 summarizes the details of all the datasets.

DL model and Preprocessing

The hyperspectral image carries hundreds of spectral reflectance bands over a wide range of wavelengths. As a result of these, the HSIs data is highly redundant. Principle component analysis is used in the experiment to reduce the redundancy of the data. More than 98 percent of the energy is carried by the first 30 PCA components of the IP dataset and the first 15 PCA components of the PV and SL datasets. So these selected PCA components are further used in the experiments.

The neighbouring pixel is highly correlated in hyperspectral images. Image patches of size 5 X 5 are extracted and given input to the CNN model to take advantage of the spatial correlation.

Roy et al. (Roy et al. 2019) proposed a Hybrid Spectral Convolutional Neural Network (HybridSN), a combination of 2D and 3D convolutional layers. A 2D—CNN could not consider the channel-related information. A 3D—CNN carries a complex model that needs high computational cost, large memory requirements, and data quality and noise sensitivity. The HybridSN has three 3D convolutional layers followed

by a 2D convolutional layer. The dense and dropout layers combine the extracted features from the convolutional layer. The HybridSN, a customized CNN model, is used for further experiments.

AL framework

AL is an iterative learning process. In every iteration, the set of samples is selected from the data pool based on the query selection techniques. These selected samples are labelled by the human annotator and added to the training dataset, and removed from the data pool. This iterative process is executed until the budget is exhausted. There are many criteria for the termination of the iterative process, like the number of iterations, total labelled samples, the accuracy achieved and execution time. This experiment uses the number of labelled samples as the budget. The necessary steps for AL are shown in Algorithm 1. A small number of samples from each class are initially labelled as D_{train} , whereas D_{pool} refers to the remaining unlabeled dataset. The D_{train} dataset trains the CNN model. The CNN model receives D_{pool} samples as input to determine the prediction uncertainty and other statistics needed by the various AL methods. The batch of samples identified as D_{select} is chosen for manual labelling based on the selection criteria of the AL technique. After providing a label to the D_{select} , samples are included in the D_{train} and removed from the unlabelled D_{pool} dataset.

```

Require: Initial  $D_{train}$ ,  $D_{pool}$ 
while (!Budget exhausted) do
  Obtain desired features of  $D_{pool}$ 
   $D_{select}$  = Selected batch of samples (n) for manual labelling
  Acquire labels of  $D_{select}$ 
   $D_{pool} = D_{pool} / D_{select}$ 
   $D_{train} = D_{train} \cup D_{select}$ 
  Retrain classifier with  $D_{train}$ 
end while

```

Algorithm 1 Active deep learning

AL methods

The following section describes the four basic AL methods implemented with the CNN classifier for HSIs classification. Assume $x_i \in D_{pool}$, is one of the unlabeled samples, k is the total number of unlabelled samples, and C is the number of different classes. $P_j(x_i)$ is the prediction probability of sample x_i of class j.

- Random Sampling (RS): RS selects a random batch of samples. The results of this method are used as a baseline.
- Least Confidence (LC): LC technique selects a batch of samples for annotation for which the classifier is least confident for the class prediction.

$$\min_{i=1..k} P_L(x_i) \quad (1)$$

Where P_L is the maximum class prediction probability of sample x_i .

- Breaking Tie (BT): The classifier was unable to distinguish between two classes, as indicated by the smaller difference between the highest and second-highest class prediction probabilities. BT chooses the samples with the smallest difference in class prediction probabilities.

$$\min_{i=1..k} P_B(x_i) - P_{SB}(x_i) \quad (2)$$

P_B and P_{SB} are sample x_i 's highest and second-highest class prediction probability, respectively.

- Entropy Sampling (ES): Higher entropy indicates the most uncertain sample for the prediction. ES selects the set of samples with the higher entropy.

$$\max_{i=1..k} - \sum_{j=1..c} P_j(x_i) \log P_j(x_i) \quad (3)$$

In multiclass classification, all class prediction probabilities are used for entropy calculation. In contrast, LC used maximum prediction probability, and BT used maximum and second maximum class prediction probability for uncertainty measures. So the ES is a reliable measure of uncertainty compared to the LC and BT.

Hyperparameters and evaluation measures

AL model training is an iterative process. During every iteration, the batch of samples (n) is added to the training set, and

the model gets retrained with the updated training dataset. This iterative training process ends with the termination criteria, the total number of labelled samples in this experiment. Here 800 labeled samples are used as the budget. In every iteration, the CNN model is retrained with five epochs.

With a learning rate of 0.001, the CNN model trains iteratively using the Softmax activation function. Patch augmentation techniques like flipping and rotation have been used to reduce the overfitting issue. Each patch from the training dataset is augmented with either flipping or rotation.

The three hyperspectral data sets under consideration are used in each experiment. With 32 training batch sizes, the Adam optimizer, and a learning rate of 0.001. The first training dataset and testing set are made before each implementation is run. Five samples from each class are taken as the initial training dataset for the experiment, and 30% of test data are used to evaluate the classifier's performance. And the remaining dataset is initialized as an unlabeled dataset D_{pool} .

Overall accuracy (OA), Average accuracy (AA), and Cohen kappa are defined to evaluate the performance of the CNN model.

Result analysis

Experiment 1 This experiment examines the effects of the CNN model train without active learning and shows the outcomes of four AL techniques. With 800 labelled samples, Table 4 shows the OA, AA and Kappa for each acquisition function. The experiment was repeated five times, and the final results were determined by averaging the outcomes from each repetition. Specifically focusing on the IP data set, the CNN classifier with ES-criterion achieves the highest overall accuracy compared to the other functions, reaching 94.63% with nearly 8% of the labelled samples. In comparison, the LC achieves the lowest OA with 79.22%. The IP dataset has two classes, Oats and Alfalfa, with only 20 and 46 total samples, respectively, even though ES achieved average class accuracy of 95%. The ES and BT AL methods also exhibit promising overall and average accuracy. In contrast, the RS and LC methods yield lower accuracy than

Table 4 Test accuracy on IP, SL and PV datasets

		CNN	RS	LC	ES	BT
IP	OA	90.35	90.11	79.22	94.63	92.26
	AA	91.85	89.38	82.82	95.00	94.04
	K- Kappa	0.89	0.89	0.77	0.94	0.91
SL	OA	94.72	94.94	92.92	98.66	96.16
	AA	95.85	95.95	96.07	99.38	97.43
	K- Kappa	0.94	0.94	0.92	0.99	0.96
PV	OA	96.37	96.33	92.96	99.21	97.25
	AA	95.25	95.15	91.28	98.93	96.46
	K- Kappa	0.96	0.95	0.91	0.99	0.96

the ES and BT methods. Notably, the performance of the RS and LC methods is even inferior to training a CNN model without any AL techniques.

The SL data set has similar behaviour, with ES and BT techniques achieving better OA than LC. With only 1.47% of the labelled samples, the ES active learning can reach 98.66% accuracy, with LC having the lowest OA at 92.92%. The ES methods often produced better OA and AA with the SL data set than the RS and LC.

Finally, in the PV image, the ES achieves 99.21% with 1.87% of the labelled samples as training dataset, which is the best OA with strong generalization capability. With 800 labeled samples, the BT and RS techniques achieved 97.25 and 96.33% of OA, respectively.

In Fig. 4, which depicts the evolution of the DL model accuracy for each acquisition function, with various D_{train} size. With PV and SL datasets, ES clearly outperforms other tested AL methods. Wherever for the IP dataset ES starts with similar accuracy to RS but falls behind it until around D_{train} size reaches 400. ES was impaired during the initial AL iteration, producing poor performance compared to the RS due to the model uncertainty. The ES method computed the entropy from the prediction probability of the samples. In the initial phase, the model might experience difficulty correctly identifying them due to its limited knowledge. Random sampling, on the other hand, covers a broader range of instances without considering informativeness. However, as more labelled samples are obtained, the performance gap diminishes, and ES can outperform random sampling by leveraging its ability to select informative samples. Regardless of D_{train} size, the LC method achieves the lowest accuracy for all three datasets. However, the BT method performs slightly better than RS with most D_{train} sizes for all three datasets.

Experiment 2 The experiment is conducted to observe the effect of batch size (n) selection for active deep learning. Here, the batch size (n) is the number of samples selected for manual labelling in every iteration. The IP dataset is used for the observation because IP has limited samples, and the number of labelled samples of different classes is also imbalanced. As shown in experiment 1, the breaking tie active learning method performs consistently on all the datasets with various iterations. So, the BT active learning

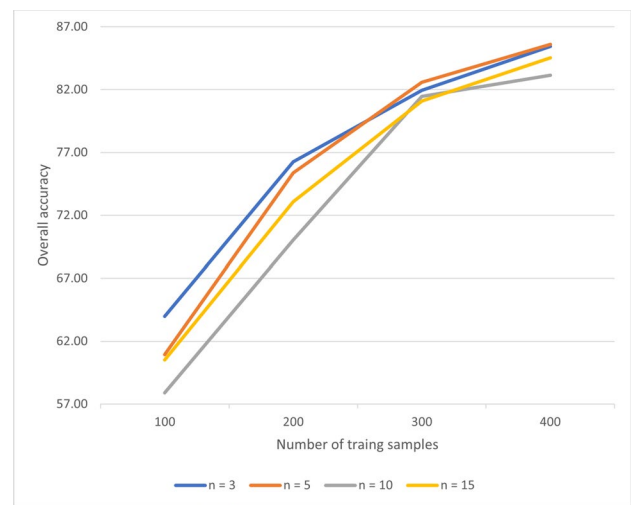


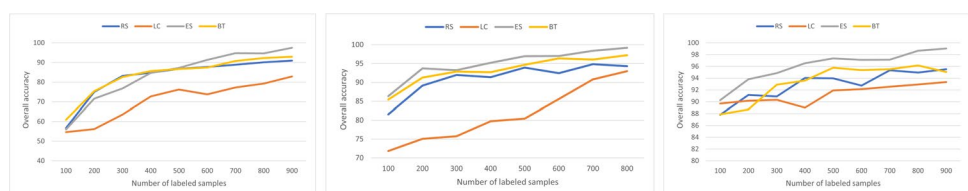
Fig. 5 Test OA on IP dataset with different batch size

method is used for the comparative analysis of different batch sizes. Figure 5 shows testing OA with batch sizes 3, 5, 10, and 15. As a result, the performance with batch sizes n=3 and 5 is better than the n=10 and 15. The batch size plays a significant role when working with a small number of labelled samples. This is evident in the graph, where the accuracy difference is significant with only 100 labelled samples, while the variations in accuracy become relatively smaller when using 400 labelled samples. Conversely, the small batch size will increase the computation load because the number of iterations will increase for the labelled samples. Based on the results, a moderate batch size of five is used for the experiment.

Conclusion and discussion

The article’s objective is to anticipate a broader area of Hyperspectral image classifications from an active learning perspective. Even though some deep learning techniques perform very well, more research is required to reach a definitive conclusion on the influence of active learning on the choice of a minimally informative set of samples for labelling in supervised learning paradigms. The results of the scientific studies show a significant improvement in accuracy rates with simple to advanced AL approaches, mainly when using DL models.

Fig. 4 Comparative analysis on IP (left), PV (middle) and SL (right) dataset



The first section introduces the fundamental theoretical ideas about the most popular AL approaches. This study examines the advantages and disadvantages of the above methods concerning the classification of hyperspectral images. The study concluded that the learning-based AL methods are gaining popularity due to their adaptability compared to the other basic AL methods like uncertainty-based, margin-based, and committee-based. However, their primary application domains are language processing, computer vision, and object detection. It is worth noting that learning-based AL techniques currently need to be widely employed in hyperspectral image classifications.

The various advancements are critically analyzed and categorized broadly to improve the AL method. Table 1 summarizes different categories and tasks adopted for active learning in the domain of the HSIs classification based on advancements in AL. The Multiple Views Active Learning (MVAL) techniques demonstrate improved performance compared to basic active learning methods. However, the inclusion of multiple views leads to increased computation costs. A cost-effective MVAL solution can enhance the classification of hyperspectral images when labelled samples are limited. By combining pixels and generating superpixels, the candidate samples selected based on these superpixels can reduce the dependency on labelled datasets. Various methods are available for generating superpixels, which can be explored and applied to HSIs.

In various computer vision applications, the advanced deep learning methods and the complex structures that can extract hidden features from data play a crucial role. Table 2 gives insights into Active deep learning approaches applied for the HSIs classification. Table 2 focuses on the task performed, AL technique, specialized features, and classifier for the Active deep learning for HSIs classifications. In the literature, most ADL methods select batch of the sample in every AL iteration. The main goal of the AL framework is to identify the most informative and diverse samples for annotation. However, when multiple samples are selected using the same criteria, a single batch may contain similar samples from the same category, which could hamper the performance of the AL algorithm as it may not receive a representative set of samples from various categories. To overcome this issue, more research is required.

Furthermore, the experimentation section uses the deep learning model CNN with the different AL methods for the comparative analysis. The result of experimentation shows the improvement in the CNN model using iterative AL techniques. The entropy-based AL technique has been observed to outperform the RS, LC and BT for IP, SL, and PV datasets. The batch size also plays a vital role in active learning. The experiment aimed to determine the efficacy of different batch sizes for active deep learning. The investigation can explore the possibility of adjusting the batch

size based on the learning stage of the DL model. Additionally, determining the optimal batch size selection can be studied to enhance the learning process of the DL algorithm during AL iterations. Adopting the batch size and optimizing its selection is expected to improve the overall performance and effectiveness of the DL model within the AL framework. Implementing AL method within the Active Deep Learning framework helps to alleviate the reliance on labelled data. Research in this area holds great potential for the remote sensing community, enabling them to effectively utilize remote sensing data for various applications. This advancement can significantly contribute to remote sensing and its diverse applications.

Acknowledgements We are grateful to the reviewer for their critical and insightful comments. Their feedback has greatly improved the quality of our work, and their dedication and expertise are highly appreciated.

Author's contribution Both the author have equally contributed..

Funding Funding not obtained for this study.

Data availability This manuscript has no associated data.

Declarations

Competing interests The authors declare no competing interests.

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

References

- Abdi G, Samadzadegan F, Reinartz P (2017) Spectral–spatial feature learning for hyperspectral imagery classification using deep stacked sparse autoencoder. *Journal of Applied Remote Sensing* 11(4):042604
- Abe N (1998) Query learning strategies using boosting and bagging. *Proceedings of the Fifteenth International Conference on Machine Learning (ICML98)* pp 1–9
- Adão T, Hruška J, Pádua L et al (2017) Hyperspectral imaging: A review on uav-based sensors, data processing and applications for agriculture and forestry. *Remote Sensing* 9(11):1110
- Ahmad M, Shabbir S, Oliva D et al (2020) Spatial-prior generalized fuzziness extreme learning machine autoencoder-based active learning for hyperspectral image classification. *Optik* 206:163712
- Ahmad M (2020) Fuzziness-based spatial-spectral class discriminant information preserving active learning for hyperspectral image classification. *arXiv preprint arXiv:200514236*
- Azadegan A, Porobic L, Ghazinoory S et al (2011) Fuzzy logic in manufacturing: A review of literature and a specialized application. *Int J Prod Econ* 132(2):258–270
- Cai R, Liu C, Li J (2021) Phase-induced gabor-based multiview active learning for hyperspectral image classification. *IEEE Geosci Remote Sens Lett* 19:1–5
- Cao X, Yao J, Fu X et al (2020a) An enhanced 3-d discrete wavelet transform for hyperspectral image classification. *IEEE Geosci Remote Sens Lett* 18(6):1104–1108

- Cao X, Yao J, Xu Z et al (2020b) Hyperspectral image classification with convolutional neural network and active learning. *IEEE Trans Geosci Remote Sens* 58(7):4604–4616
- Chen Y, Jiang H, Li C et al (2016) Deep feature extraction and classification of hyperspectral images based on convolutional neural networks. *IEEE Trans Geosci Remote Sens* 54(10):6232–6251
- Crawford MM, Tuia D, Yang HL (2013) Active learning: Any value for classification of remotely sensed data? *Proc IEEE* 101(3):593–608
- Culotta A, McCallum A (2005) Reducing labeling effort for structured prediction tasks. In: *AAAI*, pp 746–751
- Datta D, Mallick PK, Bhoi AK et al (2022) Hyperspectral image classification: Potentials, challenges, and future directions. *Comput Intell Neurosci* 2022:1–36
- Demir B, Persello C, Bruzzone L (2010) Batch-mode active-learning methods for the interactive classification of remote sensing images. *IEEE Trans Geosci Remote Sens* 49(3):1014–1031
- Deng L, Yu D et al (2014) Deep learning: methods and applications. *Foundations and Trends in Signal Processing* 7(3–4):197–387
- Deng C, Xue Y, Liu X et al (2018) Active transfer learning network: A unified deep joint spectral–spatial feature learning model for hyperspectral image classification. *IEEE Trans Geosci Remote Sens* 57(3):1741–1754
- Di W, Crawford MM (2010) Multi-view adaptive disagreement based active learning for hyperspectral image classification. In: 2010 *IEEE International Geoscience and Remote Sensing Symposium*, *IEEE*, pp 1374–1377
- Di W, Crawford MM (2011) View generation for multiview maximum disagreement based active learning for hyperspectral image classification. *IEEE Trans Geosci Remote Sens* 50(5):1942–1954
- Ding C, Zheng M, Chen F et al (2022) Hyperspectral image classification promotion using clustering inspired active learning. *Remote Sensing* 14(3):596
- Dong S, Quan Y, Feng W et al (2021) A pixel cluster cnn and spectral-spatial fusion algorithm for hyperspectral image classification with smallsize training samples. *IEEE J Sel Top Appl Earth Obs Remote Sens* 14:4101–4114
- Fang B, Li Y, Zhang H et al (2020) Collaborative learning of lightweight convolutional neural network and deep clustering for hyperspectral image semi-supervised classification with limited training samples. *ISPRS J Photogramm Remote Sens* 161:164–178
- Feng Y, Yuan Y, Lu X (2017) Learning deep event models for crowd anomaly detection. *Neurocomputing* 219:548–556
- Gao AF, Rasmussen B, Kulits P, et al (2021) Generalized unsupervised clustering of hyperspectral images of geological targets in the near infrared. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pp 4294–4303
- Guo J, Zhou X, Li J et al (2016) Superpixel-based active learning and online feature importance learning for hyperspectral image analysis. *IEEE J Sel Top Appl Earth Obs Remote Sens* 10(1):347–359
- Hajiani F, Parhizgar N, Keshavarz A (2021) Hyperspectral image classification using cluster based graph regularized low rank representation and dictionary learning. *Neurocomputing* 462:208–220
- Haußmann M, Hamprecht FA, Kandemir M (2019) Deep active learning with adaptive acquisition. *arXiv preprint arXiv:1906.11471*
- Haut JM, Paoletti ME, Plaza J et al (2018) Active learning with convolutional neural networks for hyperspectral image classification using a new bayesian approach. *IEEE Trans Geosci Remote Sens* 56(11):6440–6461
- He Z, Liu H, Wang Y et al (2017) Generative adversarial networks-based semisupervised learning for hyperspectral image classification. *Remote Sensing* 9(10):1042
- He X, Chen Y, Ghamisi P (2019) Heterogeneous transfer learning for hyperspectral image classification based on convolutional neural network. *IEEE Trans Geosci Remote Sens* 58(5):3246–3263
- Houlsby N, Huszár F, Ghahramani Z, et al (2011) Bayesian active learning for classification and preference learning. *arXiv preprint arXiv:1112.5745*
- Hsu WN, Lin HT (2015) Active learning by learning. In: *Twenty-Ninth AAAI conference on artificial intelligence*
- Hu J, He Z, Li J et al (2018) 3d-gabor inspired multiview active learning for spectral-spatial hyperspectral image classification. *Remote Sensing* 10(7):1070
- Jamshidpour N, Safari A, Homayouni S (2020) A ga-based multi-view, multilearner active learning framework for hyperspectral image classification. *Remote Sensing* 12(2):297
- Jia S, Jiang S, Lin Z et al (2021) A survey: Deep learning for hyperspectral image classification with few labeled samples. *Neurocomputing* 448:179–204
- Joshi AJ, Porikli F, Papanikolopoulos N (2009) Multi-class active learning for image classification. In: 2009 *IEEE conference on computer vision and pattern recognition*, *IEEE*, pp 2372–2379
- Kahraman S, Bacher R (2021) A comprehensive review of hyperspectral data fusion with lidar and sar data. *Annu Rev Control* 51:236–253
- Kalita I, Kumar RNS, Roy M (2021) Deep learning-based cross-sensor domain adaptation under active learning for land cover classification. *IEEE Geosci Remote Sens Lett* 19:1–5
- Khan MJ, Khan HS, Yousaf A et al (2018) Modern trends in hyperspectral image analysis: A review. *Ieee Access* 6:14118–14129
- Konyushkova K, Sznitman R, Fua P (2017) Learning active learning from data. *Advances in Neural Information Processing Systems* 30
- Lee H, Kwon H (2017) Going deeper with contextual cnn for hyperspectral image classification. *IEEE Trans Image Process* 26(10):4843–4855
- Lei Z, Zeng Y, Liu P et al (2021) Active deep learning for hyperspectral image classification with uncertainty learning. *IEEE Geosci Remote Sens Lett* 19:1–5
- Leskes B (2005) The value of agreement, a new boosting algorithm. In: Auer Peter, Meir Ron (eds) *International conference on computational learning theory*. Springer, Berlin, pp 95–110
- Lewis DD, Catlett J (1994) Heterogeneous uncertainty sampling for supervised learning. In: *Machine learning proceedings 1994*. Elsevier, p 148–156
- Li Z, Liu M, Chen Y et al (2021) Deep cross-domain few-shot learning for hyperspectral image classification. *IEEE Trans Geosci Remote Sens* 60:1–18
- Li J (2015) Active learning for hyperspectral image classification with a stacked autoencoders based neural network. In: 2015 *7th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, *IEEE*, pp 1–4
- Liang X, Zhang Y, Zhang J (2021) Attention multisource fusion-based deep few-shot learning for hyperspectral image classification. *IEEE J Sel Top Appl Earth Obs Remote Sens* 14:8773–8788
- Lin J, Zhao L, Li S et al (2018) Active-learning-incorporated deep transfer learning for hyperspectral image classification. *IEEE J Sel Top Appl Earth Obs Remote Sens* 11(11):4048–4062
- Liu P, Zhang H, Eom KB (2016) Active deep learning for classification of hyperspectral images. *IEEE J Sel Top Appl Earth Obs Remote Sens* 10(2):712–724
- Liu C, He L, Li Z et al (2017) Feature-driven active learning for hyperspectral image classification. *IEEE Trans Geosci Remote Sens* 56(1):341–354
- Liu B, Yu X, Yu A et al (2018a) Spectral-spatial classification of hyperspectral imagery based on recurrent neural networks. *Remote Sens Lett* 9(12):1118–1127
- Liu B, Yu X, Yu A et al (2018b) Deep few-shot learning for hyperspectral image classification. *IEEE Trans Geosci Remote Sens* 57(4):2290–2304

- Liu C, Li J, He L (2018c) Superpixel-based semisupervised active learning for hyperspectral image classification. *IEEE J Sel Top Appl Earth Obs Remote Sens* 12(1):357–370
- Liu X, Hu Q, Cai Y et al (2020) Extreme learning machine-based ensemble transfer learning for hyperspectral image classification. *IEEE J Sel Top Appl Earth Obs Remote Sens* 13:3892–3902
- Lu T, Li S, Benediktsson JA (2017) Iterative clustering based active learning for hyperspectral image classification. In: 2017 IEEE international geoscience and remote sensing symposium (IGARSS), IEEE, pp 3664–3667
- Lu Q, Wei L (2021) Multiscale superpixel-based active learning for hyperspectral image classification. *IEEE Geosci Remote Sens Lett* 19:1–5
- Luo T, Kramer K, Goldgof DB, et al (2005) Active learning to recognize multiple types of plankton. *J Mach Learn Res* 6(4)
- Ma L, Ma A, Ju C et al (2016) Graph-based semi-supervised learning for spectral-spatial hyperspectral image classification. *Pattern Recognit Lett* 83:133–142
- MacKay DJ (1992) Information-based objective functions for active data selection. *Neural Comput* 4(4):590–604
- Melgani F, Bruzzone L (2004) Classification of hyperspectral remote sensing images with support vector machines. *IEEE Trans Geosci Remote Sens* 42(8):1778–1790
- Mnih V, Kavukcuoglu K, Silver D, et al (2013) Playing atari with deep reinforcement learning. *arXiv preprint arXiv:1312.5602*
- Mou L, Ghamisi P, Zhu XX (2017) Deep recurrent neural networks for hyperspectral image classification. *IEEE Trans Geosci Remote Sens* 55(7):3639–3655
- Moughal T (2013) Hyperspectral image classification using support vector machine. *J Phys Conf Ser* 439:012042 (**IOP Publishing**)
- Mu C, Liu J, Liu Y et al (2020) Hyperspectral image classification based on active learning and spectral-spatial feature fusion using spatial coordinates. *IEEE Access* 8:6768–6781
- Murphy JM (2020) Spatially regularized active diffusion learning for highdimensional images. *Pattern Recognit Lett* 135:213–220
- Muslea I, Minton S, Knoblock CA (2006) Active learning with multiple views. *J Artif Intell Res* 27:203–233
- Paoletti ME, Haut JM, Plaza J, et al (2020) Training capsnets via active learning for hyperspectral image classification. In: *IGARSS 2020–2020 IEEE International Geoscience and Remote Sensing Symposium*, IEEE, pp 40–43
- Pasolli E, Melgani F, Bazi Y (2010) Support vector machine active learning through significance space construction. *IEEE Geosci Remote Sens Lett* 8(3):431–435
- Patel U, Dave H, Patel V (2021) Hyperspectral image classification using uncertainty and diversity based active learning. *Scalable Comput Pract Exper* 22(3):283–293
- Patel U, Dave H, Patel V (2020) Hyperspectral image classification using semisupervised learning with label propagation. In: 2020 IEEE India Geoscience and Remote Sensing Symposium (InGARSS), IEEE, pp 205–208
- Priya T, Prasad S, Wu H (2015) Superpixels for spatially reinforced bayesian classification of hyperspectral images. *IEEE Geosci Remote Sens Lett* 12(5):1071–1075
- Qian Y, Ye M, Zhou J (2012) Hyperspectral image classification based on structured sparse logistic regression and three-dimensional wavelet texture features. *IEEE Trans Geosci Remote Sens* 51(4):2276–2291
- Qin Y, Bruzzone L, Li B et al (2019) Cross-domain collaborative learning via cluster canonical correlation analysis and random walker for hyperspectral image classification. *IEEE Trans Geosci Remote Sens* 57(6):3952–3966
- Ren P, Xiao Y, Chang X et al (2021) A survey of deep active learning. *ACM Comput Surv (CSUR)* 54(9):1–40
- Rottmann M, Kahl K, Gottschalk H (2018) Deep bayesian active semisupervised learning. In: 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA), IEEE, pp 158–164
- Roy SK, Krishna G, Dubey SR et al (2019) Hybridsn: Exploring 3-d–2-d cnn feature hierarchy for hyperspectral image classification. *IEEE Geosci Remote Sens Lett* 17(2):277–281
- Sabour S, Frosst N, Hinton GE (2017) Dynamic routing between capsules. *Advances in Neural Information Processing Systems* 30
- Salem MB, Ettabaa KS, Bouhleb MS (2016) Hyperspectral image feature selection for the fuzzy c-means spatial and spectral clustering. In: 2016 International Image Processing, Applications and Systems (IPAS), IEEE, pp 1–5
- Samat A, Li J, Liu S et al (2016) Improved hyperspectral image classification by active learning using pre-designed mixed pixels. *Pattern Recogn* 51:43–58
- Samat A, Li J, Lin C et al (2019) Edge gradient-based active learning for hyperspectral image classification. *IEEE Geosci Remote Sens Lett* 17(9):1588–1592
- 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(2):243–248
- Sener O, Savarese S (2017) Active learning for convolutional neural networks: A core-set approach. *arXiv preprint arXiv:1708.00489*
- Seung HS, Opper M, Sompolinsky H (1992) Query by committee. In: *Proceedings of the fifth annual workshop on Computational learning theory*, pp 287–294
- Stutz D, Hermans A, Leibe B (2018) Superpixels: An evaluation of the state-of-the-art. *Comput vis Image Underst* 166:1–27
- Sun Y, Li J, Wang W, et al (2016) Active learning based autoencoder for hyperspectral imagery classification. In: 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), IEEE, pp 469–472
- Thoreau R, Achard V, Risser L et al (2022) Active learning for hyperspectral image classification: A comparative review. *IEEE Geosci Remote Sens Mag* 10(3):256–278
- Tran T, Do TT, Reid I, et al (2019) Bayesian generative active deep learning. In: *International Conference on Machine Learning*, PMLR, pp 6295–6304
- Tuia D, Ratle F, Pacifici F et al (2009) Active learning methods for remote sensing image classification. *IEEE Trans Geosci Remote Sens* 47(7):2218–2232
- Tuia D, Volpi M, Copa L et al (2011) A survey of active learning algorithms for supervised remote sensing image classification. *IEEE J Sel Top Signal Process* 5(3):606–617
- Tuia D, Muñoz-Marí J, Camps-Valls G (2012) Remote sensing image segmentation by active queries. *Pattern Recogn* 45(6):2180–2192
- Wang G, Ren P (2020) Hyperspectral image classification with feature-oriented adversarial active learning. *Remote Sens* 12(23):3879
- Wang J, Guo S, Huang R et al (2021a) Dual-channel capsule generation adversarial network for hyperspectral image classification. *IEEE Trans Geosci Remote Sens* 60:1–16
- Wang X, Tan K, Pan C et al (2022) Active deep feature extraction for hyperspectral image classification based on adversarial learning. *IEEE Geosci Remote Sens Lett* 19:1–5
- Wang X, Xing C, Feng Y, et al (2021b) A novel hyperspectral image change detection framework based on 3d-wavelet domain active convolutional neural network. In: 2021b IEEE International Geoscience and Remote Sensing Symposium IGARSS, IEEE, pp 4332–4335
- Wu H, Prasad S (2017) Semi-supervised deep learning using pseudo labels for hyperspectral image classification. *IEEE Trans Image Process* 27(3):1259–1270

- Xing C, Ma L, Yang X (2016) Stacked denoise autoencoder based feature extraction and classification for hyperspectral images. *J Sens* 2016:1–10
- Xu X, Li J, Li S (2017) Multiview intensity-based active learning for hyperspectral image classification. *IEEE Trans Geosci Remote Sens* 56(2):669–680
- Xu M, Zhao Q, Jia S (2021) Multiview spatial-spectral active learning for hyperspectral image classification. *IEEE Trans Geosci Remote Sens* 60:1–15
- Xue Z, Zhou S, Zhao P (2018) Active learning improved by neighborhoods and superpixels for hyperspectral image classification. *IEEE Geosci Remote Sens Lett* 15(3):469–473
- Yang L, MacEachren AM, Mitra P et al (2018) Visually-enabled active deep learning for (geo) text and image classification: a review. *ISPRS Int J Geo Inf* 7(2):65
- Zhang Z, Pasolli E, Crawford MM et al (2015) An active learning framework for hyperspectral image classification using hierarchical segmentation. *IEEE J Sel Top Appl Earth Obs Remote Sens* 9(2):640–654
- Zhang F, Du B, Zhang L et al (2016a) Hierarchical feature learning with dropout k-means for hyperspectral image classification. *Neurocomputing* 187:75–82
- Zhang X, Song Q, Gao Z et al (2016b) Spectral–spatial feature learning using cluster-based group sparse coding for hyperspectral image classification. *IEEE J Sel Top Appl Earth Obs Remote Sens* 9(9):4142–4159
- Zhang H, Li Y, Zhang Y et al (2017) Spectral-spatial classification of hyperspectral imagery using a dual-channel convolutional neural network. *Remote Sens Lett* 8(5):438–447
- Zhang Y, Cao G, Li X et al (2018) Cascaded random forest for hyperspectral image classification. *IEEE J Sel Topics Appl Earth Obs Remote Sens* 11(4):1082–1094
- Zhao W, Chen X, Bo Y et al (2019) Semisupervised hyperspectral image classification with cluster-based conditional generative adversarial net. *IEEE Geosci Remote Sens Lett* 17(3):539–543
- Zhong Y, Ma A, Soon Ong Y et al (2018) Computational intelligence in optical remote sensing image processing. *Appl Soft Comput* 64(75):93
- Zhou X, Prasad S, Crawford MM (2016) Wavelet-domain multi-view active learning for spatial-spectral hyperspectral image classification. *IEEE J Sel Top Appl Earth Obs Remote Sens* 9(9):4047–4059
- Zhu L, Chen Y, Ghamisi P et al (2018) Generative adversarial networks for hyperspectral image classification. *IEEE Trans Geosci Remote Sens* 56(9):5046–5063
- Zhu JJ, Bento J (2017) Generative adversarial active learning. arXiv preprint arXiv:170207956

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.