Contents lists available at ScienceDirect

International Journal of Applied Earth Observation and Geoinformation

Large-scale land use/land cover extraction from Landsat imagery using feature relationships matrix based deep-shallow learning

Peng Dou [a], Huanfeng Shen [b,*], Chunlin Huang [a,*], Zhiwei Li [c,d], Yujun Mao [b], Xinghua Li [e]

[a] Key Laboratory of Remote Sensing of Gansu Province, Heihe Remote Sensing Experimental Research Station, Northwest Institute of Eco-Environment and Resources, Chinese Academy of Sciences, Lanzhou, China
[b] School of Resource and Environmental Sciences, Wuhan University, China
[c] JC STEM Lab of Earth Observations, Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Hung Hom, Hong Kong
[d] Research Centre for Artificial Intelligence in Geomatics, The Hong Kong Polytechnic University, Hung Hom, Hong Kong
[e] School of Remote Sensing and Information Engineering, Wuhan University, China

ARTICLE INFO

ABSTRACT

Deep learning has demonstrated its effectiveness in capturing high-level features, with convolutional neural networks (CNNs) excelling in remote sensing classification. However, CNNs encounter challenges when applied to Landsat images with limited multi-spectral bands, as they struggle to learn stable features from the spectral domain and integrate them with spatial features to enhance accuracy. Additionally, most CNN applications focus on learning features directly from the raw image, making them susceptible to spectral environment changes. To overcome these limitations, this paper introduces a novel approach for large-scale Land Use/Land Cover (LULC) extraction from Landsat OLI images. The proposed classification architecture comprises two modules. The first module utilizes a feature relationships matrix to generate an extent feature map (EFM), and a specifically designed CNN structure learns deep features from the EFM and spatial domain. In the second module, a multiple classifiers system (MCS) is employed to obtain shallow learning features, which are further enhanced by another CNN structure through continued learning. The combined features from these modules contribute to improved classification of remote sensing images. Experimental results demonstrate that our proposed method effectively acquires stable features for training deep learning models with strong generalization ability. It exhibits significant advancements in accuracy improvement and large-scale LULC extraction in the Yangtze River Economic Belt (YREB) in China, surpassing comparative approaches based on deep learning and non-deep learning methods.

1. Introduction

In the past four decades, Landsat has been widely used for large-scale land use and land cover (LULC) extraction in numerous studies (Phiri et al., 2018; Homer et al., 2020). Freely available Landsat-based LULC products with a 30 m resolution, including Finer Resolution Observation and Monitoring of Global Land Cover, GlobeLand30, and Normalized Urban Areas Composite Index, have been developed and widely applied in various scientific researches (Yu et al., 2013; Brovelli et al., 2015; Li et al., 2021). However, most of these products rely solely on original spectral features or manually selected features as inputs for the model, lacking an automatic feature learning process during remote sensing (RS) image classification, resulting in significant manual and time costs for data production. Additionally, traditional machine learning methods

have certain limitations, such as the requirement for a large number of samples and the complexity of model training, all of which impact the classification accuracy of remote sensing images classification (Du et al., 2020). This poses certain challenges for the extraction of large-scale LULC from Landsat images.

In recent years, the field of RS image classification has seen significant progress with the widespread adoption and development of deep learning technology (Davydzenka et al., 2022; Li et al., 2024). Deep learning models autonomously learn deep features from data, leading to improved classification accuracy. Various deep neural networks, including convolutional neural networks (CNNs), have been effectively used for LULC extraction (Cheng et al., 2022; Dou et al., 2024). All of them are better equipped for handling large-scale datasets and demonstrate stronger generalization and portability compared to traditional

* Corresponding authors.
E-mail addresses: shenhf@whu.edu.cn (H. Shen), huangcl@lzb.ac.cn (C. Huang).

methods (DeLancey et al., 2019; Zhang et al., 2021a). As a result, CNNs have gained considerable attention in Landsat image classification, showing excellent performance (Xu et al., 2018b; Zhang et al., 2018; Zhang et al., 2022a).

For RS image classification, CNNs can be broadly categorized into scene classification-based CNNs and semantic segmentation-based CNNs (Zhao et al.,2019; Xu et al., 2022; Maggiolo et al.,2022). Scene classification-based CNNs typically consist of convolutional and pooling layers, followed by a fully connected layer connected to a softmax layer that provides the predicted class distribution. These models take batches of scenes as inputs and yield class labels corresponding to the center points of the scenes. Initially used in Landsat image classification, they struggle with accurately identifying object boundaries due to limitations in scene size and receptive field, and are also time-consuming (Sharma et al., 2018; Zhao et al.,2019). Semantic segmentation-based CNNs offer solutions to these issues through end-to-end architectures consisting of fully convolutional layers, up or down sampling layers, skip structures, and other components (Zhao et al., 2019). Several semantic segmentation-based CNNs, including Unet, Pyramid scene parsing network, fully convolutional network, and VGG-16 network (Zhang et al., 2018; Mohith and Karthi, 2022), have been utilized for LULC extraction from Landsat RS images. However, it is worth noting that due to the 30 m spatial resolution of Landsat images, some of these models may lose detailed information. To address this, Zhao et al. (2019) proposed the entropy and Markov random field model-CNN to reduce pixel uncertainty in Landsat images while preserving detailed information, achieving high-quality LULC extraction at a large scale. However, the authors utilized the Global Land Cover 30, which has an overall accuracy (OA) of 80 %, as labels for the training set. Further evaluation is needed to assess the impact of the remaining 20 % error on classification accuracy.

Many CNN-based methods have primarily focused on learning spatial features from RS images. However, RS images are characterized by multispectral features, highlighting the need to incorporate spectral information into CNN-based methods to achieve spectral-spatial learning and enhance RS image classification. Recent attention has been given to CNN-based spectral-spatial methods, particularly in hyperspectral image classification, leading to the development of excellent models such as three-dimensional CNN, spectral-spatial residual network, and S^3Net (Paoletti et al., 2018; Zhong et al., 2018; Zhang et al., 2021b). Hyperspectral imagery provides rich spectral information for deep feature extraction. However, extracting deep features from multispectral images with relatively fewer bands presents a challenge. Consequently, researchers are actively exploring the utilization of combined spatial and spectral features for LULC extraction in multispectral imagery. For instance, permuted local spectral-spatial convolutional network has been employed for PolSAR-Multispectral data (Gopal Singh et al., 2020), while spectral-spatial convolution neural networks have been utilized for object-oriented RS image classification (Cui et al., 2018), among others. However, the development of CNN-based spectral-spatial models for large-scale LULC extraction from Landsat images remains scarce due to the limited number of bands and resolution constraints. Further research is needed to explore how to implement spectral-spatial deep learning methods for Landsat image classification.

Continuous learning and building upon existing knowledge have proven to be effective in improving classification accuracy. Drawing from this principle, Dou et al. proposed the deep-shallow learning (DSL) method in 2021 (Dou et al., 2021a). This approach leverages multiple shallow classifiers to initially classify the image, and then combines the classification results with the original data as new features for deep learning classification. DSL benefits from ensemble learning (Dou et al., 2021a; Abdali et al., 2024), overcoming the limitations posed by the scarcity of bands in RS images, and has demonstrated excellent performance in classifying multispectral RS images, including Landsat images. However, the original DSL model focuses primarily on pixel-level

classification and lacks spatial features. By incorporating CNNs-based models known for their ability to extract deep spatial features, it is anticipated that the classification performance of DSL can be further improved while building upon its original foundations.

Based on the above analysis, learning spectral-spatial features from Landsat images and continuing with deep learning on the basis of shallow classification results, is expected to yield stable features. In this paper, we propose a new method called the feature relationships matrix-based deep-shallow learning (FRM_DSL) for large-scale LULC extraction from Landsat images. This approach combines spectral-spatial CNNs with the DSL framework. Initially, we employ a feature relationship matrix (FRM) to create an extended feature map (EFM) and devise a CNN architecture for spectral-spatial feature extraction. Furthermore, we integrate a shallow learning module to improve classification performance of the FRM_DSL model. This novel approach successfully extracts detailed LULC information from Landsat OLI images on a large scale.

2. Study area and data

2.1. Study area

In order to gather a sufficient number of samples to extract representative land use features for large-scale LULC classification research, this paper selects the Yangtze River Economic Belt (YREB) in China as study area (Fig. 1). The YREB spans across China from east to west, characterized by complex terrain and diverse climate variations. The average annual temperature ranges from 14.2 to 17.8 °C, with precipitation levels varying between 873.9 and 2397.5 mm. The YREB holds great significance for China. It covers 21.3 % of China's total land area (approximately 73,634.75 km^2) and supports over 40 % of the country's population and gross domestic product (Xu et al., 2018a; Han et al., 2021).

2.2. Data description

A total of 308 Landsat 8 Operational Land Imagery (OLI) images, acquired in 2019 and 2020 and with a cloud cover of less than 5 %, were downloaded from the US Geological Survey website. For the classification purposes of this study, we exclusively employed bands 1 to 7 with a 30-meter resolution. Subsequently, using ENVI software and high-resolution images from Google Earth, experts conducted a comprehensive analysis of all the images and meticulously selected evenly distributed regions of interest (ROIs) with typical features as samples. The classification schedule and number of sample ROIs are outlined in Table 1.

Accurately distinguishing fallow land and bare soil based solely on spectral information from the Landsat OLI posed a challenge. Therefore, the bare land in this study includes fallow land. Following the sampling process, experts cross-checked each other's work to ensure high-quality ROIs for the samples. Subsequently, all the images were clipped into tiles with a pixel size of 64 × 64, discarding any tiles that did not contain ROIs. Finally, the ROIs were utilized to assign labels to each tile, producing datasets for spring, summer, autumn and winter.

3. Methods

3.1. Space-spectrum joint classification based on feature relationship matrix

Space-spectrum joint deep learning approaches have shown promising results in enhancing the accuracy of RS image classification (Pan et al., 2019; Gao et al., 2020). However, when dealing with Landsat images, which have only a few multi-spectral bands, learning deep features solely from the spectral domain becomes challenging, and subsequently merging them with deep spatial features to improve
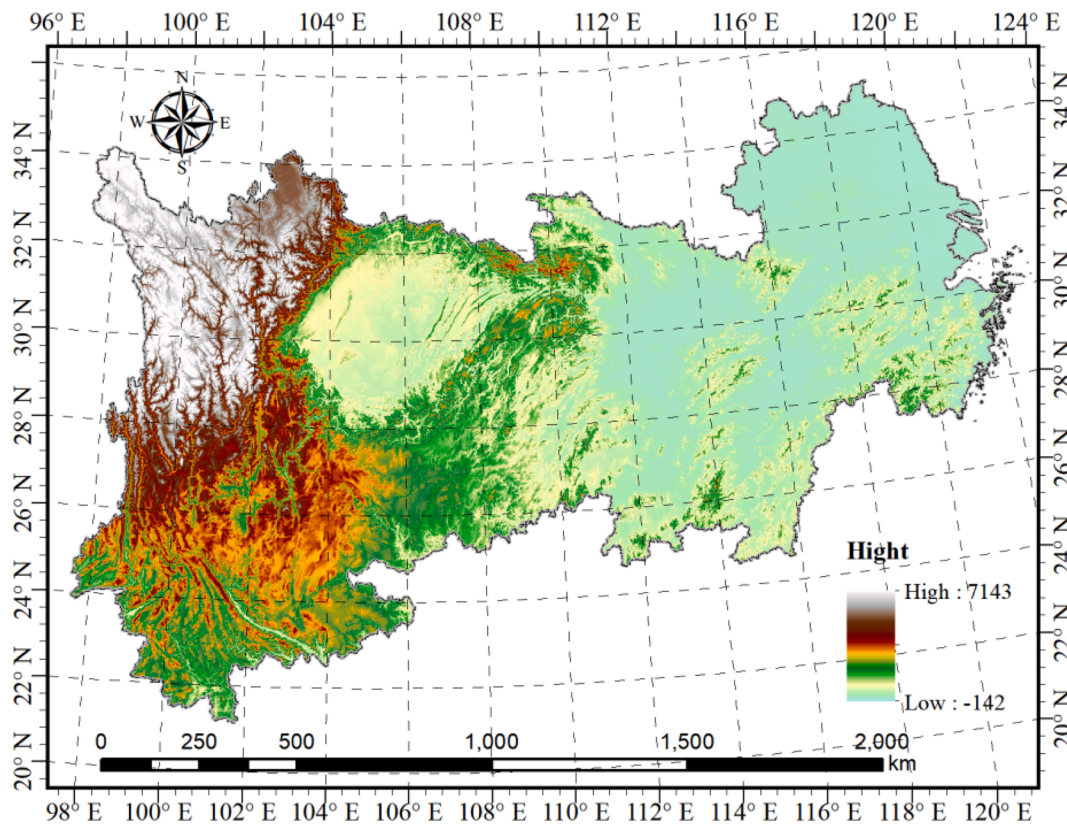
Fig. 1. Study area.

Table 1
Class schedule and number of sample ROIs for this study.

Class	Label	Number of sample ROIs				Total number
		Spring	Summer	Fall	Winter	
Cropland	CL	1199	679	1331	1682	4891
Forest	FO	1773	1036	3160	2210	8179
Grassland	GR	329	542	1030	799	2700
Shrub	SH	1118	1140	107	1591	3956
Wetland	WL	524	523	1722	740	3509
Water body	WB	4276	1103	2703	3982	12,064
Impervious surface	IS	1779	905	1994	2026	6704
Bare land	BL	1806	1048	1969	3373	8196
Ice and snow	IAS	298	88	756	765	1907
Total Number of sample ROIs		13,102	7064	14,772	17,168	52,106
Total number of tiles		6540	4203	9057	8543	28,343

classification accuracy poses additional difficulties (Pasquarella et al., 2018). In order to overcome these challenges, we utilized the spectral bands to establish a FRM to obtain an EFM. Our approach incorporated CNN structures designed to learn deep features from both the spatial and spectral domains. The proposed model consists of two key modules: the spatial feature learning module and the EFM learning module. A visual representation of the model is illustrated in Fig. 2.

3.1.1. Extended feature map with use of feature relationship matrix

The limited number of spectral bands in Landsat images can have a negative impact on the accuracy of LULC extraction. To overcome this limitation, a novel approach involves mapping features derived from the spectral bands into a two-dimensional domain and subsequently using CNNs to extract deep features (Dou and Zeng, 2020). In this study, we employ the FRM to establish an EFM, which maps the enhanced features of the Landsat OLI image into a two-dimensional domain. Firstly, each

pixel in an image is represented as a vector, with its entries corresponding to the spectral features in each band. The value of each pair of entries is then calculated using expression (1) as follows.

$$V_{ij} = a \times \mathrm{band}_i + b \times \mathrm{band}_j \tag{1}$$

where $i = 1, 2,…, n; j = 1, 2,…, n; n$ is the band count of a Landsat image; a and b are two mutually unequal constants and their default values are 0.25 and 0.75, respectively.

The OLI images utilized in this paper consist of 7 bands, resulting in a 7×7 FRM matrix for each pixel. In order to obtain feature maps of the same size as the input tiles after multi-layer convolution, we padded the FRMs with the average value of the 7 bands to expand their size to 8×8. Subsequently, we rearranged all the FRMs in the order of pixels in the tile, generating a new feature map with a size of $(8 \times 64) \times (8 \times 64)$. This transformation allowed us to convert all the information from the FRMs into a two-dimensional plane, making it more convenient for the network to learn deep features using a CNN architecture. We referred to this new feature map as EFM.

3.1.2. Convolutional neural networks for classification

It is evident that the features learned from the EFM are pixel-wise and lack spatial information. This limitation hampers the ability of the EFM-based framework to achieve higher accuracy. To address this issue, we have designed a dedicated module to extract spatial features from the original image as shown in Fig. 2. This module consists of three residual network (ResNet) blocks, with each convolutional layer utilizing a 2×2 filter size and a stride of 1 pixel. We have specifically selected a 2×2 filter size over others to primarily extract more detailed features from EFM and image space while avoiding the inconvenience caused by feature blurring. Unlike traditional ResNet architectures, our module adopts a fully convolutional design, and the output feature layer preserves the same size as the input patch.

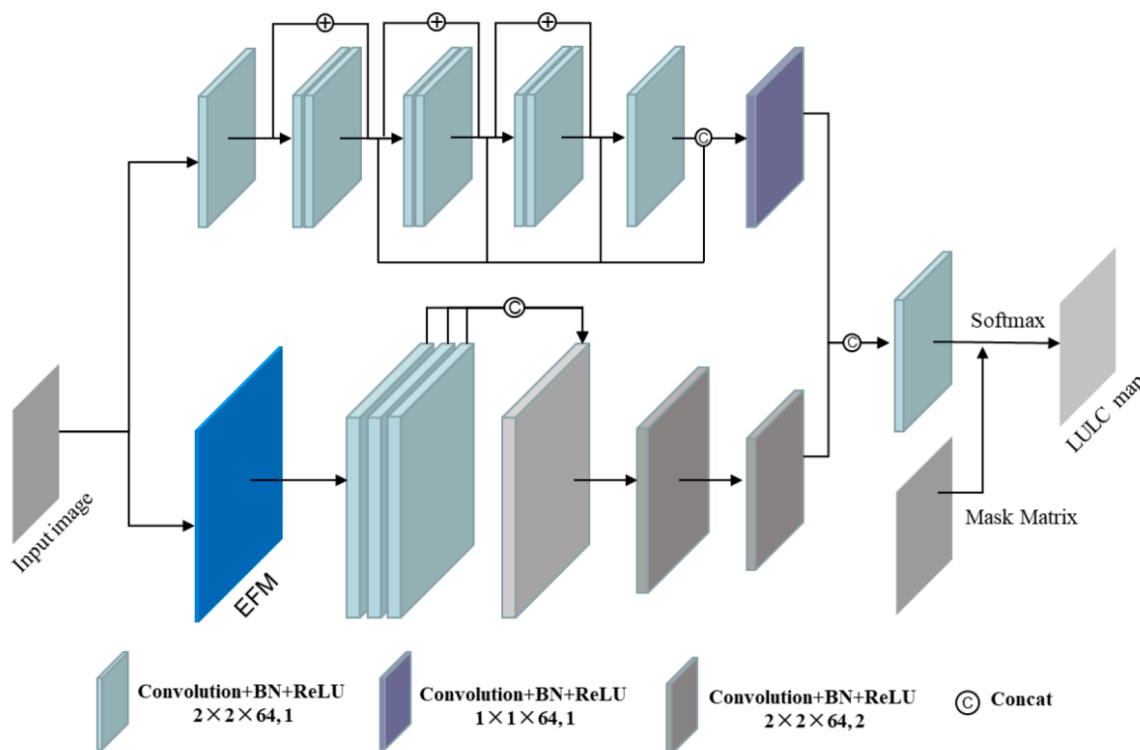Subsequently, the output layer of the ResNet module and the output

Fig. 2. Sketch map of Space-spectrum joint classification based on feature relationship matrix (SSJC_ERM) model.

layer of the EFM module are merged together using a merging layer (Fig. 2). This is followed by another convolutional layer and a softmax function to produce the final classification map. Due to the lower resolution and higher presence of mixed pixels in Landsat images compared to high-resolution remote sensing images, it is crucial to maintain the image details. Having excessively deep convolutional layers could negatively impact the recognition of object boundaries. Hence, this study employs ResNet blocks to retain detailed information during the learning process of deep spectral features.

In traditional RS image semantic segmentation tasks, all pixels in the training set are typically labeled. However, when it comes to large-scale classification, collecting such a comprehensive training set becomes extremely challenging. In our research, the training data for classification is highly sparse. Therefore, we need to compute the cross-entropy loss only with the pixels that are present in the training sample ROIs, while ignoring the pixels that are outside the ROIs. To facilitate this, we utilize a mask matrix to assist in the calculation of the loss, as proposed by Xu et al. (2020).

Let Y be the ground truth matrix with size of 64×64 in the training set, the mask matrix M can be defined as equation (2).

$$M(\mu,\nu) = \begin{cases} 1, & \text{if } Y(\mu,\nu) > 0 \\ 0, & \text{otherwise} \end{cases} \qquad (2)$$

M denotes whether pixel (μ,ν) belongs to the training set ($M(\mu,\nu) = 1$) or not. The M is utilized during the network stage, and the optimization of the network can be accomplished with the RMSprop optimizer and categorical-cross-entropy loss function (Tieleman and Hinton, 2012). In this paper, the model is fit for 500 training epochs, and the learning rate set as 0.001.

3.2. Feature relationships matrix based deep-shallow learning

The outcomes obtained through shallow learning have been recognized as valuable feature information that can enhance the classification accuracy of deep learning (Dou et al., 2021a). To further enhance the

classification performance, we have integrated the concept of DSL into the SSJC_ERM framework, resulting in FRM_DSL. The process of remote sensing image classification using FRM_DSL is illustrated in Fig. 3.

Firstly, the SSJC_ERM network is employed to learn the joint space-spectrum features from the input image. Subsequently, a multiple classifiers system (MCS) is utilized to classify the input image, yielding a series of soft classification results. These soft classification results represent the predicted probabilities of different classes by each base classifier. Next, the soft classification results are input into the convolutional layers, generating feature maps. These feature maps are then multiplied with the feature maps derived from the SSJC_ERM module, producing new feature maps. In order to retain the spatial characteristics, the new feature maps are concatenated with the feature maps derived from the SSJC_ERM module. Finally, the concatenated feature maps pass through a Softmax layer to generate a LULC map for the input image.

To facilitate the training process of shallow classifiers within the MCS, we employed the Classification and Regression Tree (CART) algorithm to implement the shallow learning module. To mitigate the impact of hard labels produced by the base classifier on the classification, we utilized the predicted probabilities for each class as soft labels. These soft labels were then employed to generate the output of the MCS. This approach helps to alleviate any potential bias or errors introduced by relying solely on hard labels for classification.

4. Experiments and results

4.1. Comparing methods

To evaluate the experimental results, three classification strategies were employed on the test data. The first strategy involved disregarding the spatial feature extraction module in the SSJC_ERM and directly utilizing the EFM feature extraction module for classification. The second strategy employed the SSJC_ERM for classification, while the third strategy utilized FRM_DSL for classification.

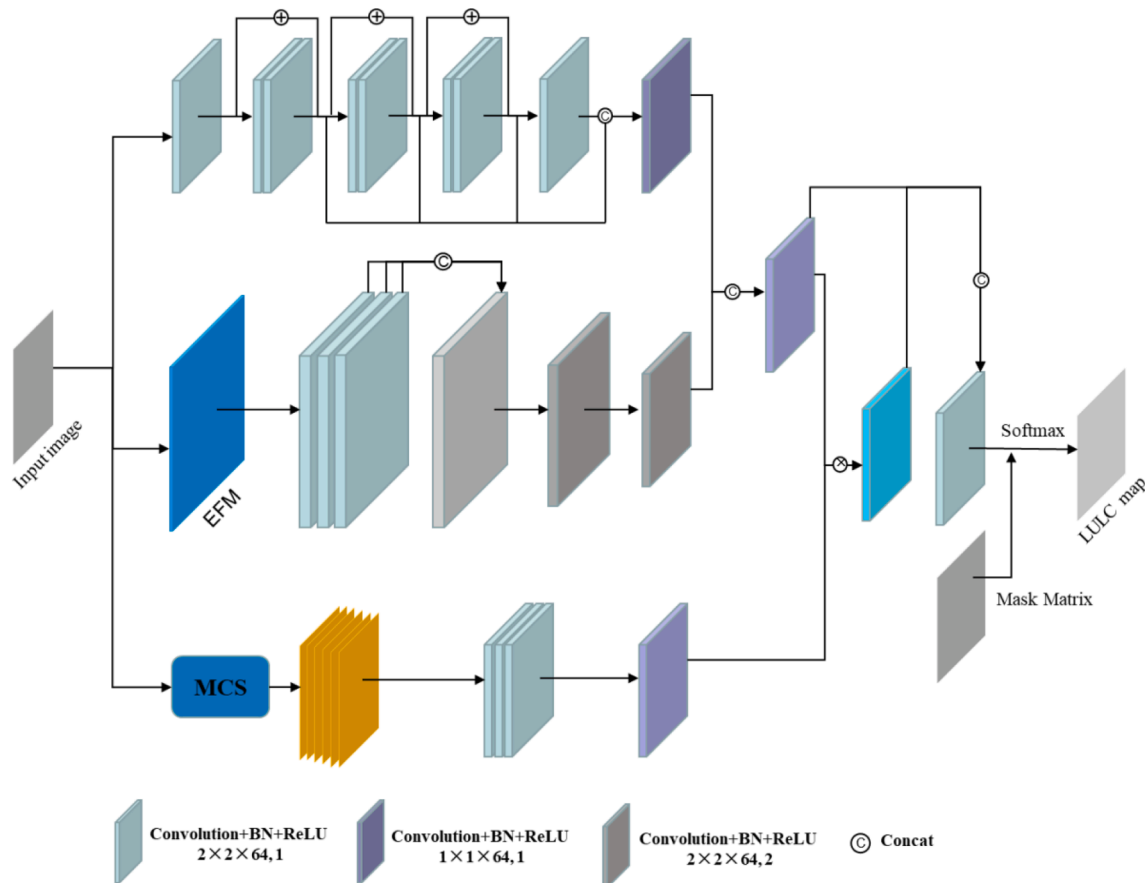Additionally, several classifiers, both deep learning and non-deep

Fig. 3. The remote sensing image classification process using the FRM_DSL.

learning, were used for comparison purposes. These classifiers included DSL (Dou et al., 2021a), Lizhiwei Net (LZWNet) (Wang et al., 2020), Spectral-Spatial Fully Convolutional Networks (SSFCN) (Xu et al., 2020), Unet, one-dimensional CNN (1D CNN) (Dou et al., 2021b), random forest (RF) (Talukdar et al., 2020), and Generative Extreme Learning Machine (GenELM). These methods have demonstrated promising performance in complex RS classification tasks and are considered representative of the current popular classification

techniques for meaningful comparison.

4.2. Accuracy evaluation

To evaluate the classification performance in this study, each dataset was randomly divided into two parts: 80 % for training and 20 % for testing. To optimize the parameters of the deep learning models, we conducted experiments with different numbers of filters, specifically 8,

Table 2
OA of the deep learning related classifiers.

Data set	Filter number	EFM	SSJC_ERM	FRM_DSL	DSL	LZWNet	And	Unet	1DCNN
Spring	8	93.90	93.91	94.46	87.68	89.70	83.39	88.27	93.90
(BS = 32)	16	94.11	94.53	94.37	86.45	91.95	88.71	**90.39**	94.11
	32	94.20	**95.39**	**95.45**	89.10	**93.69**	91.00	89.94	94.20
	64	**94.35**	94.91	94.46	**90.16**	92.63	**91.39**	89.76	**94.35**
Summer	8	84.69	85.05	87.98	77.53	68.51	49.30	74.32	47.41
(BS = 32)	16	83.28	85.04	89.22	84.15	77.70	72.26	**74.65**	78.03
	32	84.44	**86.57**	**89.53**	87.52	78.82	84.51	74.21	82.59
	64	**86.33**	85.22	87.15	86.12	77.26	**84.61**	73.33	**87.71**
Autumn	8	75.16	76.39	80.26	75.52	78.93	69.11	60.59	74.52
(BS = 32)	16	77.91	78.23	81.65	74.95	75.11	80.11	59.82	76.98
	32	**79.44**	75.14	82.47	78.14	**81.96**	79.92	61.08	76.45
	64	73.12	**79.33**	**82.56**	**79.34**	80.12	**80.12**	**62.33**	**81.35**
Winter	8	78.05	82.34	86.16	76.22	58.33	38.53	62.96	73.17
(BS = 32)	16	81.45	86.32	88.30	78.96	**77.42**	86.01	**65.60**	71.16
	32	83.28	86.62	**89.82**	81.44	78.78	86.35	63.31	73.11
	64	**84.29**	**86.76**	89.13	**82.33**	74.94	**86.70**	65.33	**77.84**
Average of max OA (BS = 32)		86.10	87.01	89.34	84.84	83.31	85.71	73.24	85.31
Average of max OA (BS = 16)		85.22	84.98	87.40	79.96	80.33	83.26	71.65	83.88
Average of max OA (BS = 64)		86.01	86.22	88.95	81.99	82.35	84.88	73.06	84.97

16, 32, and 64, and batch sizes (BSs) of 16, 32, and 64. The number of epochs was set to 500, and the snapshot with the highest accuracy was considered as the final classification accuracy. After thorough testing, it was determined that a batch size of 32 yielded the best results among these experiments. To ensure a fair comparison, we standardized the batch size for all models to 32. The corresponding overall accuracies (OAs) for different filter numbers are shown in Table 2. Finally, we select the classifiers with the best results for comparative and analysis, and the classification results at the per-class level are depicted in Fig. 4.

When considering the average maximum OA of the classifiers across the four datasets, it can be observed that the proposed method generally outperformed the deep learning methods used for comparison. The mean values of the maximum OA achieved by EFM, SSJC_ERM, and FRM_DSL on the four datasets were 86.10 %, 87.01 %, and 89.34 %, respectively. In contrast, SSFCN and 1DCNN attained high accuracies of 85.71 % and 85.31 %, respectively. On the other hand, Unet exhibited the lowest accuracy, reaching only 73.24 %. This could be attributed to the fact that Unet is designed with a deeper structure for image semantic segmentation. However, in this experiment, training was only conducted using a portion of the pixels, which resulted in a lower classification accuracy.

The division of the dataset in the experiment enables us to assess the performance of classification models. However, it should be noted that the testing data is not completely independent of the training data, and instances belonging to the same image may exhibit a high level of correlation. This can make it challenging to accurately gauge the robustness and portability of the classification models. To address this concern, we employed the classification models trained on one season's dataset to classify the testing datasets of the other three seasons. This approach allowed us to evaluate the robustness and portability of the models. The classification accuracy results are presented in Table 3, providing insights into the performance of the models across different seasons.

To enhance clarity, we have categorized the accuracy of the classifier when evaluated on validation data from the same season as local accuracy (LA), while the accuracy when evaluated on validation data from a different season is referred to as non-local accuracy (NLA). As depicted in Table 3, substantial differences emerge in the accuracy of these classifiers when utilizing samples from different seasons. For instance, the LA for the spring dataset demonstrates a high accuracy of over 88 %. However, when these classifiers were tested on samples from other seasons, the NLA was lower than the LA. This discrepancy is primarily attributed to the temporal variation in spectral features exhibited by certain objects in RS images, which is not adequately represented in the samples collected for the experiment.

In the experiment, our expectation was that a classifier would exhibit a high NLA, at least comparable to the LA. To evaluate this, we calculated the differences between LA and NLA, as illustrated in Fig. 5. It can be observed that our proposed methods exhibit smaller differences between LA and NLA compared to most of the comparative methods. This indicates that our proposed methods not only achieve high classification accuracy but also possess strong generalization capabilities.
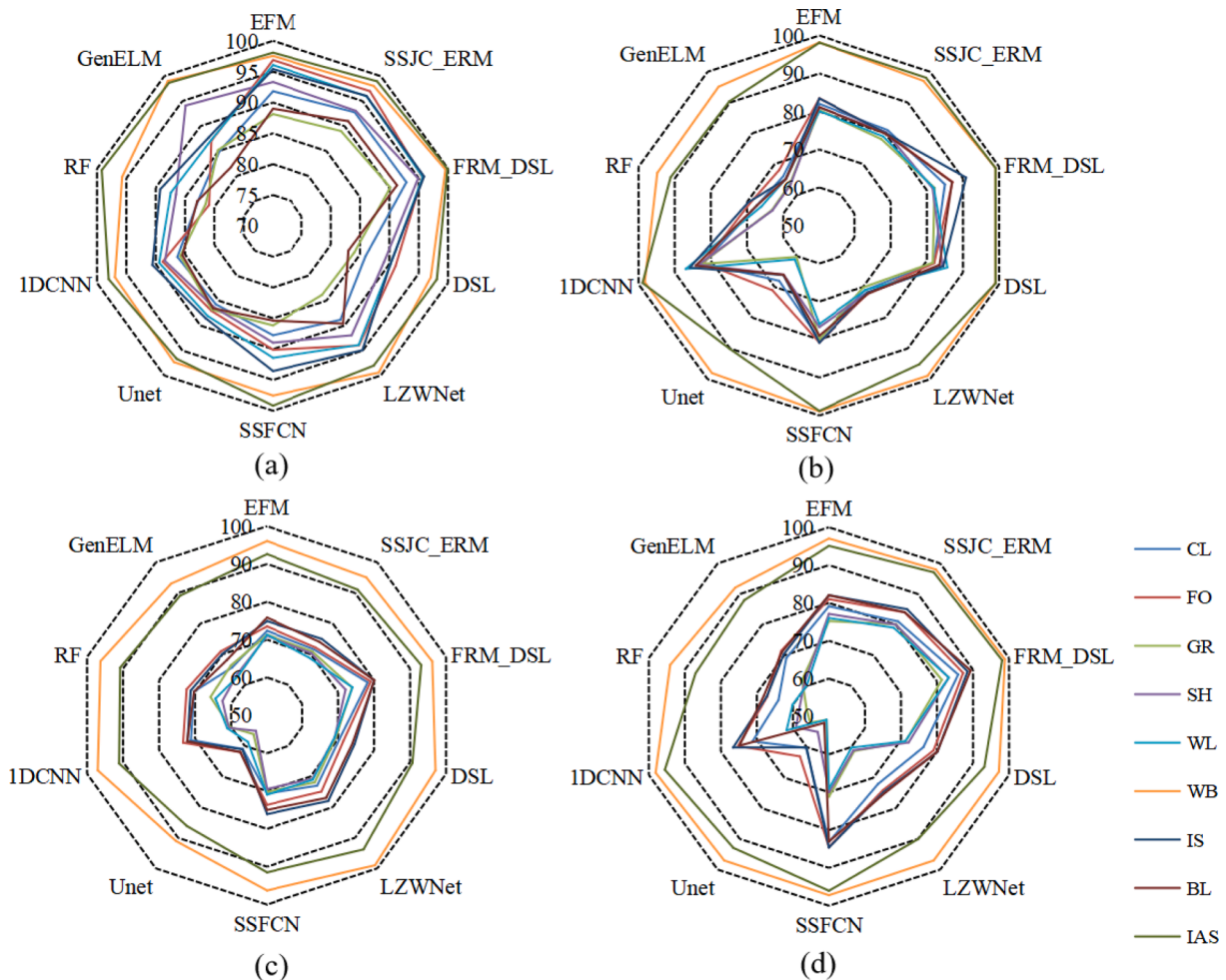

Fig. 4. Final classification accuracy at the per-class level (%): (a), (b), (c), and (d) represent the evaluations carried out using the datasets for spring, summer, autumn, and winter, respectively.

Table 3
Evaluation of classification accuracy of various classifiers across different seasons (%).

Methods	Classifier trained by spring dataset				Classifier trained by summer dataset			
	Spring	Summer	Fall	Winter	Spring	Summer	Fall	Winter
EFM	94.35	80.29	76.56	79.22	85.37	86.33	77.52	76.56
SSJC_ERM	95.39	80.02	75.29	78.51	85.69	84.57	76.39	80.21
FRM_DSL	95.45	85.55	81.00	81.30	88.26	89.53	80.24	82.56
DSL	90.16	83.27	78.26	79.91	80.12	87.52	70.56	72.36
LZWNet	93.69	73.71	75.89	79.14	74.06	78.82	66.59	60.22
SSFCN	91.39	81.86	74.84	80.97	90.91	84.61	76.57	70.80
Unet	89.76	49.12	40.54	63.18	58.07	74.65	65.97	65.33
1DCNN	90.35	83.87	67.38	79.91	68.19	87.71	73.60	77.36
RF	88.62	70.29	60.12	66.32	69.19	74.40	65.32	60.29
GenELM	89.24	71.24	57.59	62.27	62.32	73.24	65.22	60.23

Methods	Classifier trained by fall dataset				Classifier trained by winter dataset			
	Spring	Summer	Fall	Winter	Spring	Summer	Fall	Winter
EFM	92.37	81.59	79.44	76.33	89.32	79.23	81.27	84.29
SSJC_ERM	94.60	83.21	78.33	78.24	93.27	80.12	81.21	86.76
FRM_DSL	96.33	86.89	82.56	80.24	95.86	83.00	85.27	89.82
DSL	87.27	70.12	78.34	75.16	83.27	72.33	76.33	82.33
LZWNet	89.74	75.53	80.96	70.92	85.55	68.11	77.63	77.42
SSFCN	94.64	81.07	79.12	78.43	91.89	78.64	84.80	84.70
Unet	59.79	57.19	68.33	56.32	53.69	51.35	56.14	67.60
1DCNN	92.58	79.25	76.46	77.25	87.82	75.11	80.72	77.84
RF	67.22	62.69	76.27	60.32	80.69	60.33	70.33	72.23
GenELM	70.95	66.89	74.32	53.69	78.22	62.26	71.24	74.36
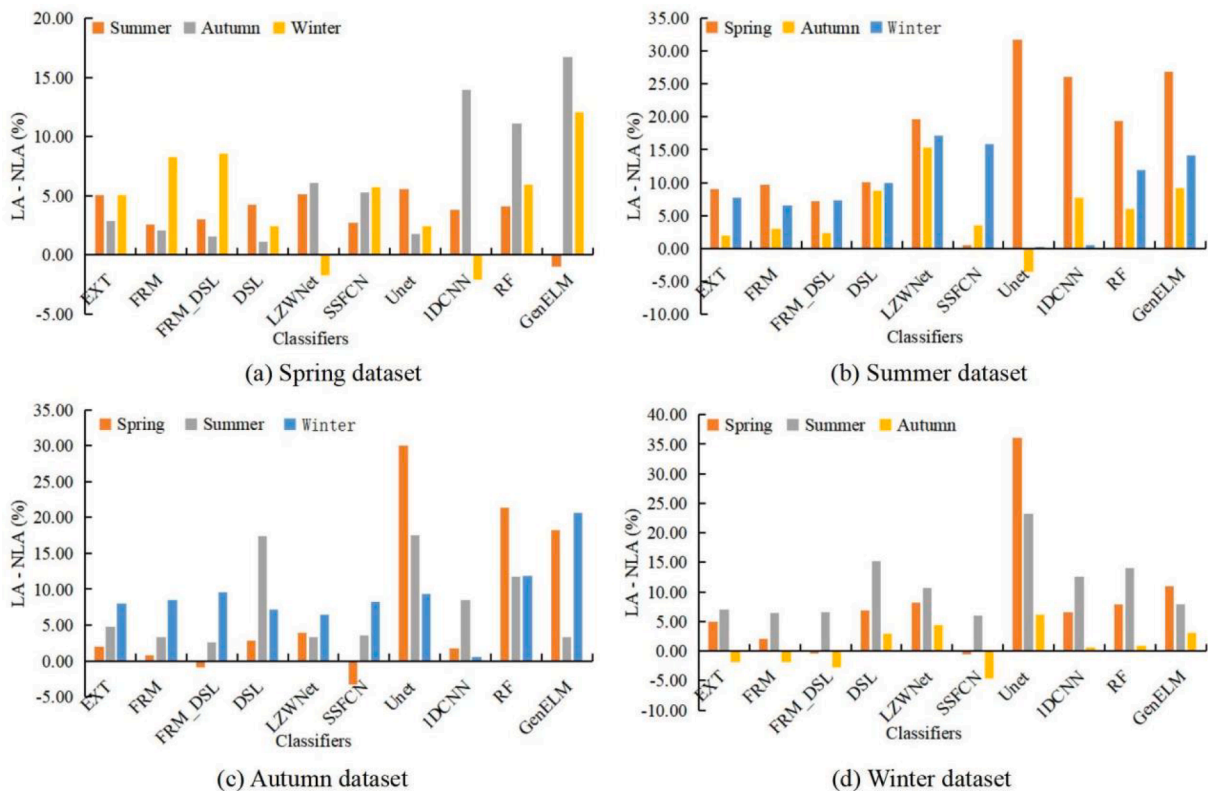
Fig. 5. The differences of LA and NLA for different classifiers trained by different datasets (the closer the percentage difference is to zero the better the model is).

4.3. LULC mapping

By implementing the proposed method, all Landsat OLI images in the collection were classified for LULC mapping across the four seasons (Fig. 6). To ensure the reliability of the LULC data, the most commonly occurring value for each pixel was calculated by analyzing the stack of all LULC maps, resulting in the final value (Fig. 7). Due to the different

acquisition times of the selected few-cloud images, certain spatially adjacent images in the western region exhibited slight boundary splicing in the seasonal LULC maps. This occurrence can be attributed to the challenging terrain and cloudy conditions, as well as the long time interval between the acquisition of two adjacent images during the process of selecting images with minimal cloud cover. In addition, classification errors also contribute to the presence of seam lines in the classified
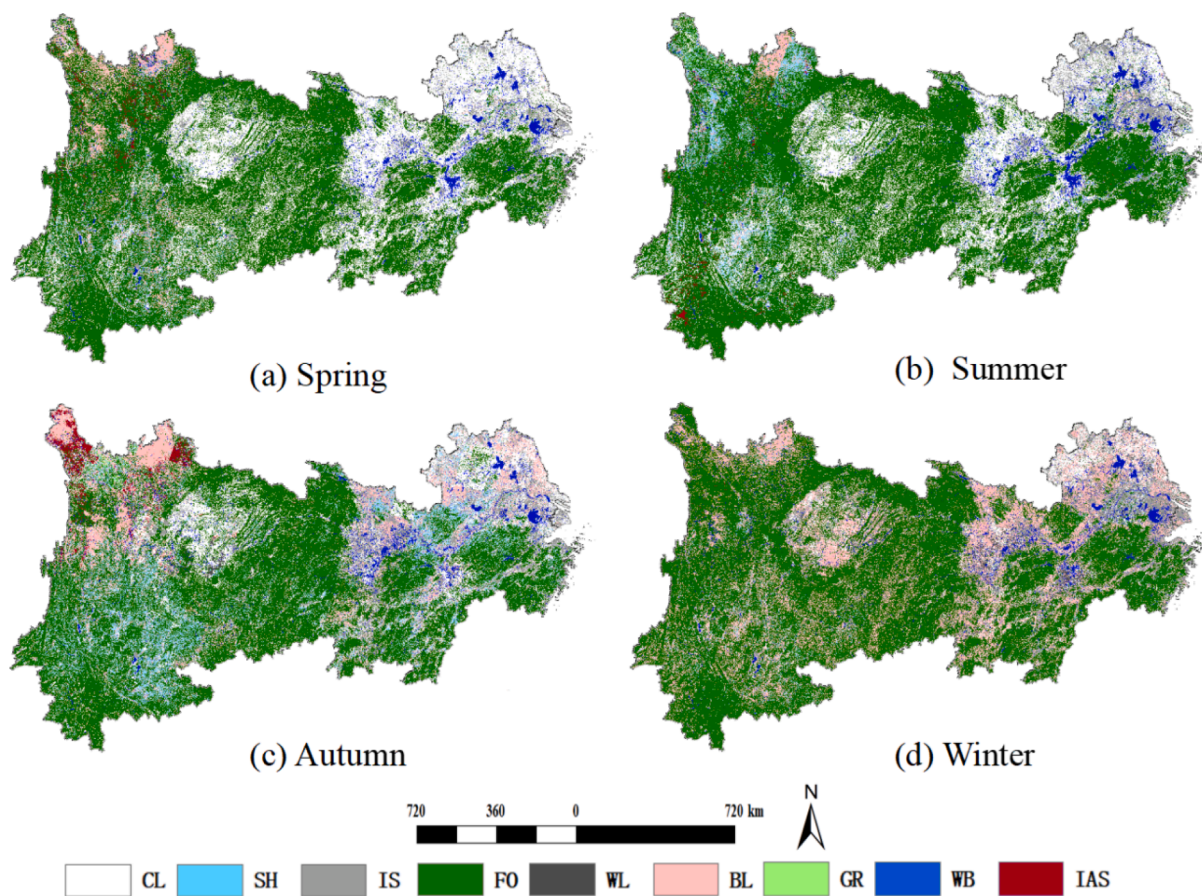
Fig. 6. LULC maps of the four seasons.

images. However, these seam lines were effectively eliminated when the seasonal LULC maps were fused into a stable LULC map.

We utilized classifiers trained on the spring dataset to classify an image acquired on April 2, 2017, along the Worldwide Reference System path 131, row 38. As depicted in Fig. 8, due to a lack of feature learning, the non-deep learning methods, RF and GenELM, exhibited poor performance. GenELM misclassified the majority of IS as WB, while RF misclassified CL as IS and bare land BL. Compared to the non-deep learning methods, the deep learning techniques demonstrated improved classification results. Especial for the proposed methods, as extracted more common features, they achieved superior LULC mapping performance compared to other approaches, and their LULC are more correspond to the actual ground objects in the RS image.

5. Discussion

5.1. Feature extraction with use of FRM_DSL

One of the advantages of EFM is its ability to provide features using expression (1) in the 2D domain, and subsequently learn deep features from them through CNNs. This approach, unlike direct utilization of CNNs on raw images, offers the benefit of preserving the intricate details of the ground objects, which is especially crucial when dealing with images at resolutions of Landsat OLI. In Fig. 9, we can observe an example of EFM: the pixels in the original image are transformed into a grid point consisting of multiple pixels in EFM. This data format visually enhances the details, facilitating feature extraction effectively.

Due to space constraints, it is not possible to display all feature maps of the models. However, we can extract the first three principal components using principal component analysis (PCA) and assess the impact of different methods using RGB color mode as depicted in Fig.10.

LZWNet and Unet, with their deeper convolution structures, fail to produce clear feature maps when learning from the original image, most likely due to the absence of strict boundary constraints within the samples investigated in this study. Additionally, the presence of check-erboard artifacts can be attributed to uneven overlap in the deconvolution process. DSL and 1DCNN directly operate on individual pixels, which leads to the presence of "salt and pepper noise" in the feature maps. Notably, the features learned by these two models, particularly for built-up areas, lack accuracy. On the other hand, SSFCN, combining spectral and spatial domain features with a shallow convolution structure, yields feature maps with relatively minimal ambiguity, although some details are still lost. EFM model excels at preserving ground details in its feature maps. By combining EFM with shallow learning features and spatial features, the objects present in the feature maps learned by FRM_DSL exhibit improved separability. Particularly, Fig. 10 (d) clearly demonstrates the significant enhancement of features in the built-up area, making them easily identifiable by the classifier compared to other methods employed, thereby improving the classification accuracy.

5.2. Impact of the training sample size on FRM_DSL

To investigate the impact of training sample size on classification accuracy, the number of training samples was increased by 10 % of the original training set. Fig. 11 (a) illustrates the results of the experiment conducted on the spring sample sets. It was observed that when the sample size was 10 %, FRM_DSL achieved an OA of 90.25 %, whereas the compared method, LZWNet, only attained an OA of 82.1 %. As the training sample size reached 100 %, the OA of FRM_DSL rose to 95.45 %, while LZWNet achieved an OA of 93.69 %. Notably, the growth rate of FRM_DSL's OA with respect to training sample size was significantly lower compared to that of LZWNet. This characteristic suggests that to
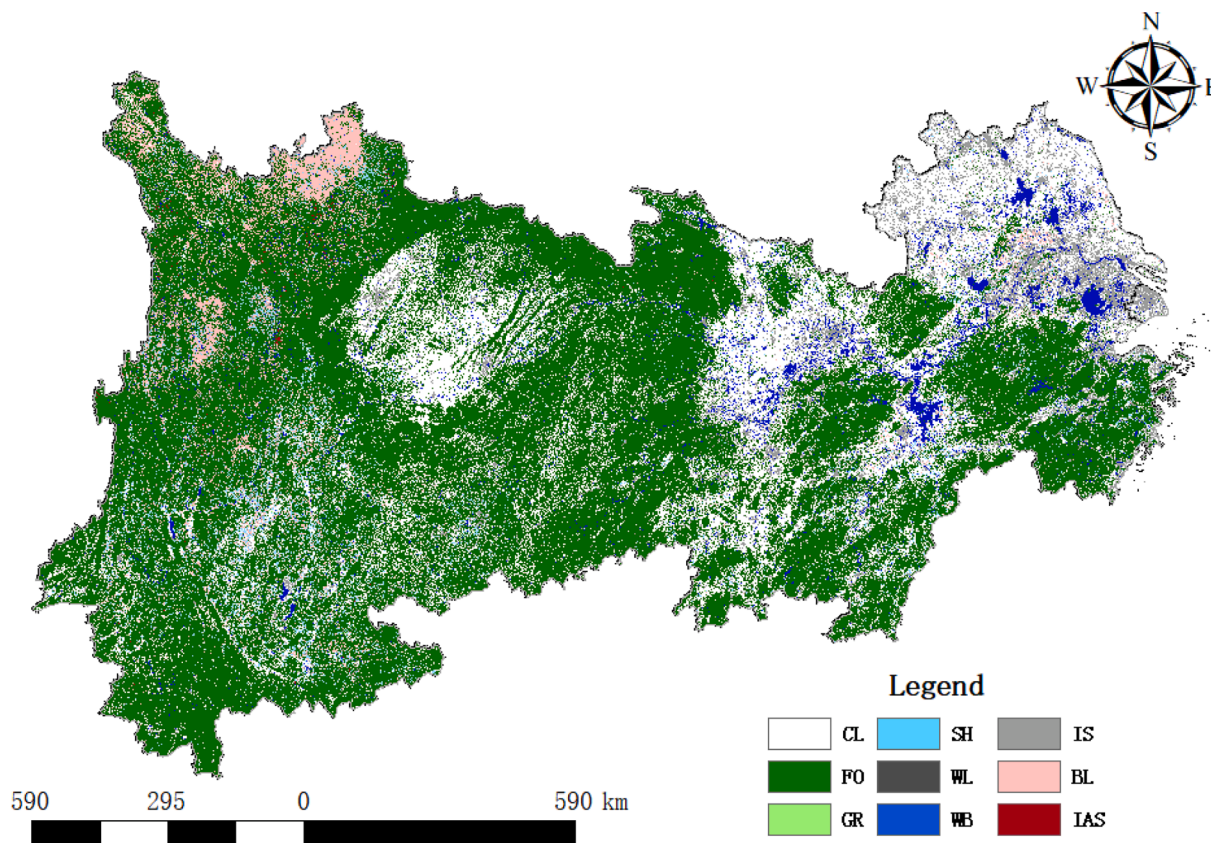
Fig. 7. Final LULC map.

achieve higher classification accuracy, LZWNet requires a larger number of samples, whereas FRM_DSL can achieve better results with fewer samples. A similar trend was observed in the experiments conducted with the summer, autumn, and winter training samples. Even with smaller sample sizes, FRM_DSL consistently exhibited higher classification accuracy compared to other methods included in the comparison. However, it is worth noting that by leveraging a larger number of samples, the accuracy of FRM_DSL can be further improved. For instance, in the experiment involving the autumn sample set, the OA of FRM_DSL demonstrated gradual improvement as the training sample size increased.

5.3. Impact of the shallow classifier on FRM_DSL

To investigate the impact of shallow classifiers on the accuracy of FRM_DSL, this study explores both the number and type of these classifiers. Fig. 12 depicts the influence of the number of shallow classifiers on the accuracy of CART-based FRM_DSL. Generally, as the number of base classifiers increases, the accuracy of FRM_DSL tends to improve. More shallow classifiers indicate higher accuracy for FRM_DSL. However, when the number of shallow classifiers is small, the accuracy of FRM_DSL can be lower than that of the classifier used for comparison, resulting in a reduction rather than an improvement in accuracy. This suggests that with a larger number of shallow classifiers, FRM_DSL can learn more stable and abstract features from the outcomes of shallow learning, thus enhancing its classification performance.

For further validation of the FRM_DSL, we trained shallow classifiers using RF, and in line with the aforementioned findings, the number of shallow classifiers in FRM_DSL was set to 50. It is worth noting that the number of CARTs plays a crucial role in RF as it directly impacts the classification accuracy (Breiman, 2001). Consequently, the number of CARTs in RF was set to 10, 20, 30, 40, and 50, respectively, to investigate the effect of shallow classifiers with varying accuracies on FRM_DSL

as shown in Table 4.

Comparing the RF-based FRM_DSL with the CART-based FRM_DSL, it is evident that the classification accuracy of the former is significantly higher. This can be attributed to the fact that a single CART is equivalent to having only one CART in RF, resulting in shallow classifiers with lower accuracy. However, as the number of CARTs in RF increases from 10 to 50, the accuracy of shallow classifiers improves, leading to further enhancements in the accuracy of the respective FRM_DSLs. This indicates that, apart from the number of shallow classifiers, the performance of FRM_DSL can also be improved by employing shallow classifiers with higher accuracy.

Furthermore, we conducted a comparison of the number of epochs required to achieve maximum accuracy for both CART-based FRM_DSL and RF-based FRM_DSL during training. Interestingly, we found that when the average accuracy of the shallow classifier is higher, RF-based FRM_DSL achieves the highest accuracy with fewer epochs. This observation suggests that the shallow learning module of FRM_DSL obtains a preliminary understanding of the target object, and utilizing these results as input to the deep learning network enables faster identification of the optimal solution during training, thereby improving convergence speed.

It is important to note that in a serial scenario, the number of shallow classifiers may increase the complexity of the model. However, the generation of each shallow classifier is an independent process, which also facilitates parallel processing. Therefore, in future research, it is necessary to optimize the model for parallel processing to further enhance operational efficiency.

5.4. LULC mapping comparison with existing products

Three scenes were selected to compare our LULC results with existing products: GlobalLand30 V2020 (https://www.globallandcover.com/), China Land Cover Dataset (CLCD) (Yang and Huang, 2021), and Global
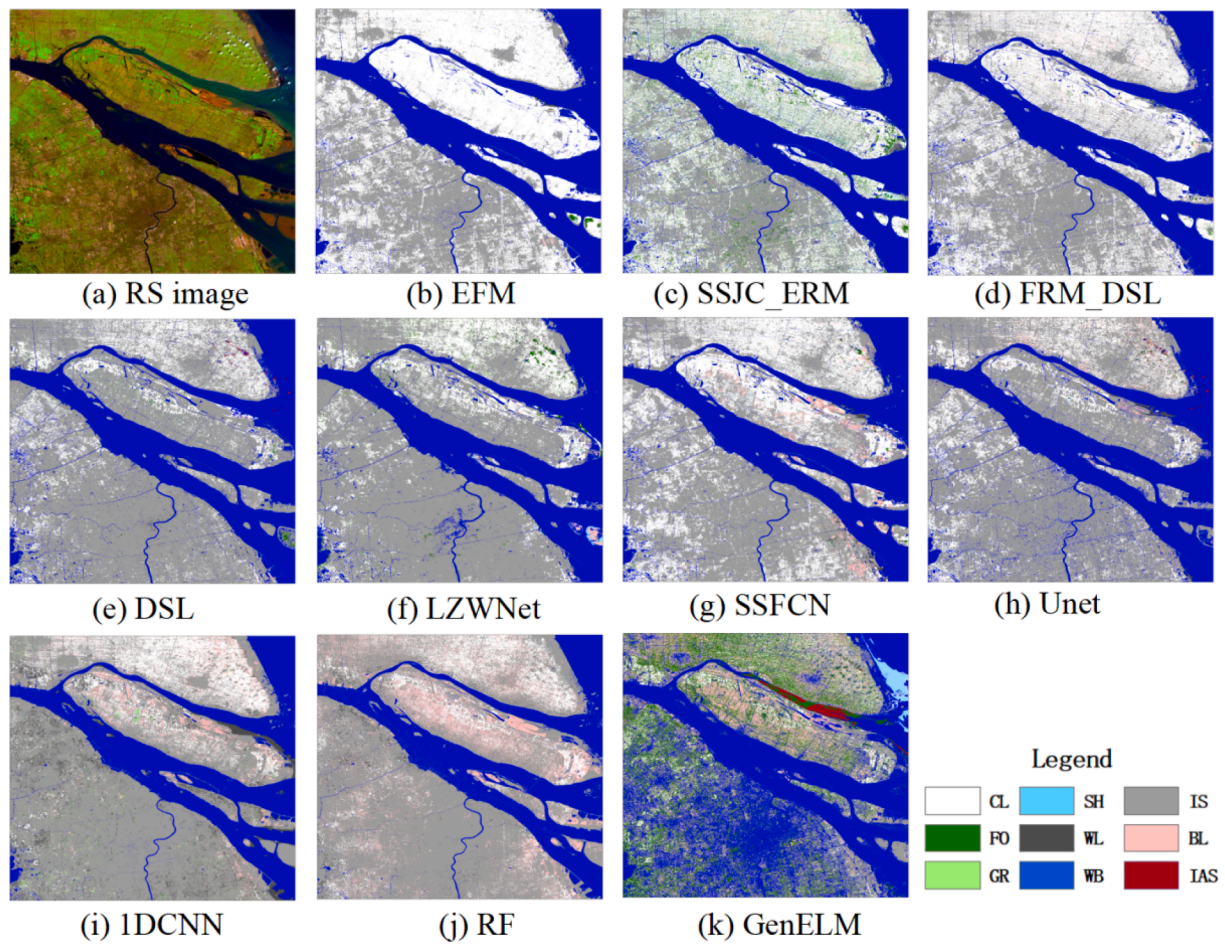
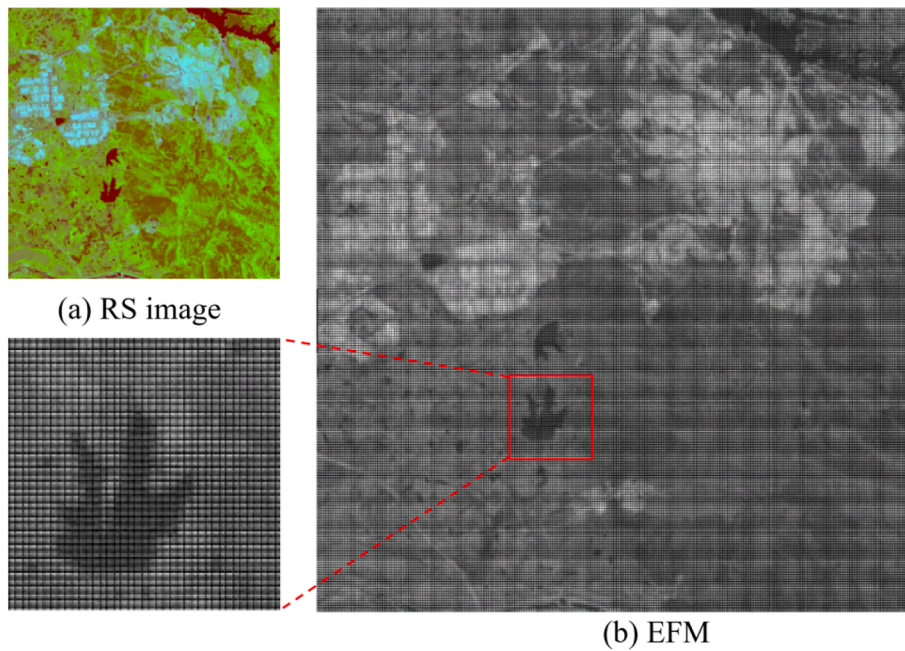**Fig. 8.** Classification results of Landsat OLI image with use of different classifiers.
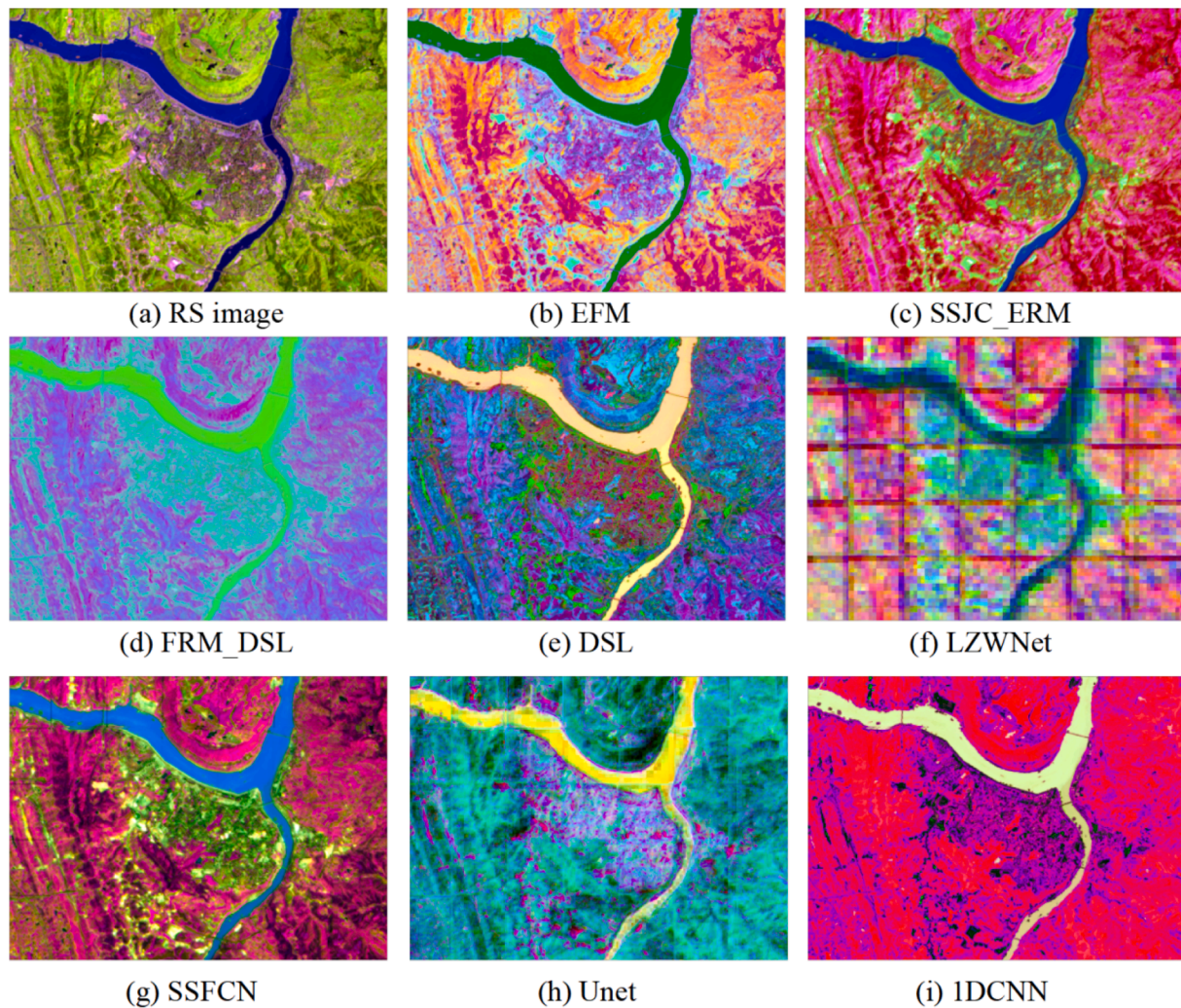

Fig. 9. The EFM of RS image.

Fig. 10. First 3 main components of feature maps derived from different models using PCA.

Land Cover with Fine Classification System at 30 m in 2020 (GLC_FCS30-2020) (Zhang et al., 2022b). These LULC products correspond to the same time period as our LULC data, representing the latest LULC products in 2020.

Fig. 13 shows the LULC mapping results for Scene 1, with several ponds observed in the regions of interest within the RS image. Upon examining the GLC_FCS30-2020, it is apparent that while the positions of these WBs were discerned to some extent, there is an issue of over-classification, leading to a lack of accurate boundary. Additionally, the under-classification of WBs in GLC_FCS30-2020, and misclassification of some as something else like CL, is evident. Similar problems are observed in the CLCD dataset. In comparison, the LULC results obtained in this study are significantly more effective in identifying and classifying WBs, with the majority being correctly labeled. Furthermore, the LULC produced in this study offers a more comprehensive representation of detailed information, such as the distribution of rivers within the urban area.

Fig. 14 shows the LULC mapping results for Scene 2 in Chengdu, China. The mountain range to the east of the city is mainly covered by forests. However, in GlobalLand30 V2020, GLC_FCS30-2020, and CLCD, much of the mountain range was misclassified as CL, with only a few areas identified as FO, making the direction of the mountain range unclear in the LULC maps. In contrast, our study's LULC results better distinguish FO and CL, accurately representing the mountain range's direction on the map. Additionally, our LULC results align well with the built-up areas observed in the RS image, while the classifications in the

other datasets are less accurate.

Fig. 15 shows the LULC mapping results in the mountainous area. However, due to the presence of shadows and variations in light intensity, the three datasets compared in this study often misclassified the shaded and sunny sides of mountains as FO and SH, respectively. Our LULC results accurately depict the distribution of SHs on the slopes and valleys without relying on the distinction between shaded and sunny sides, indicating higher accuracy in this particular area. Additionally, the methodology presented in this paper successfully extracted almost all WBs, while the CLCD dataset struggles to distinguish rivers accurately and fails to identify the lake located in the middle region, similar to GLC_FCS30.

In addition to assessing the LULC mapping effects, we also conducted evaluations of the accuracy for different products. Initially, we randomly selected 100 points from each category within our LULC product. Subsequently, in conjunction with reference images, we compared and obtained the actual labels for each point as well as the LULC labels for various LULC products. These results were then utilized to assess accuracy. Recognizing the stochastic nature of the samples, we repeated this procedure 10 times and calculated the average, resulting in an accuracy evaluation presented in the Table 5. The data reveals that our LULC exhibits a commendable overall accuracy of 84.77 %, surpassing CLCD with an accuracy of 83.49 %. Following closely are GLC_FCS30-2020 and GlobalLand30 V2020 with respective accuracies of 80.03 % and 76.89 %. Nonetheless, different LULC products demonstrate their unique classification advantages for various land cover features, which
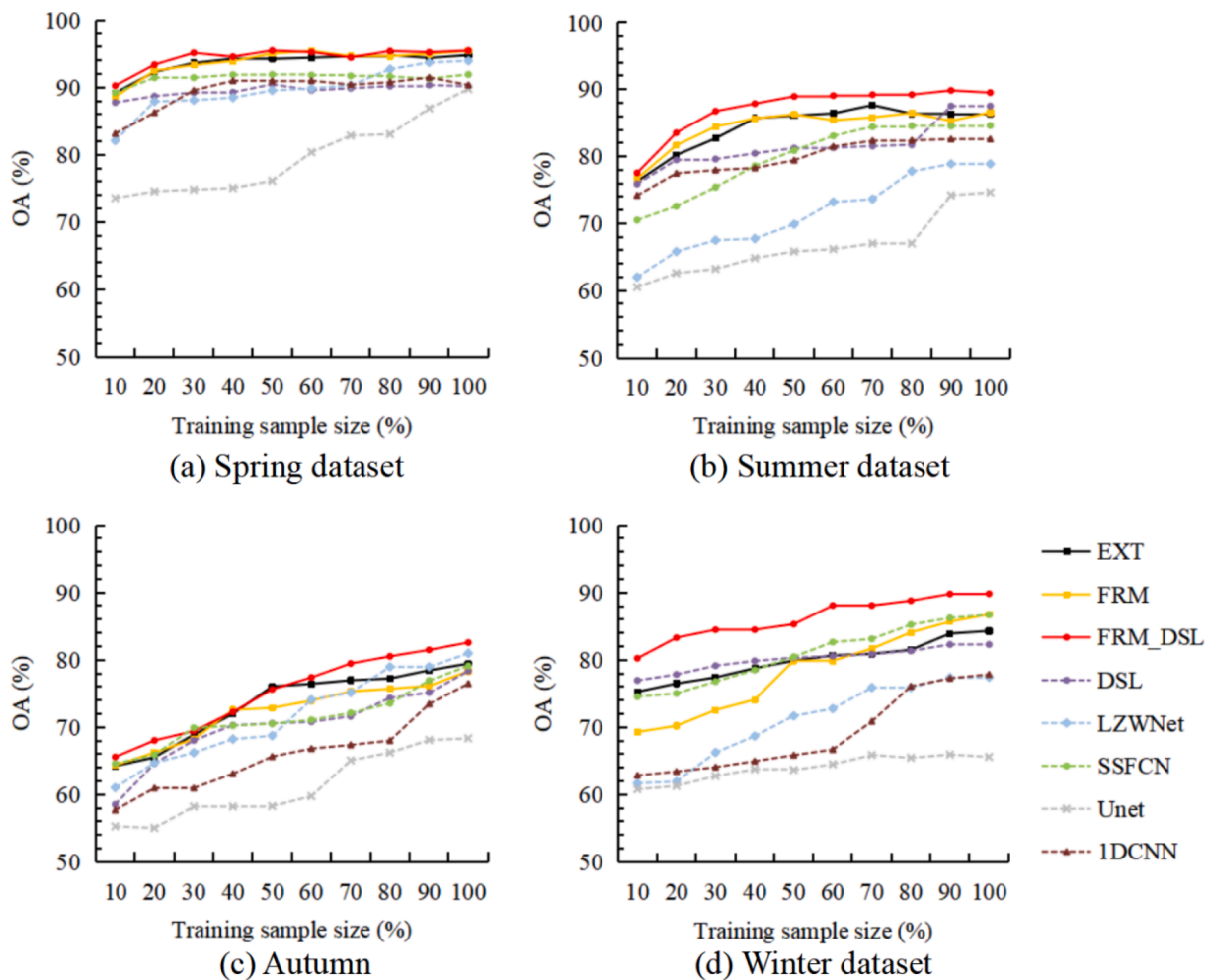
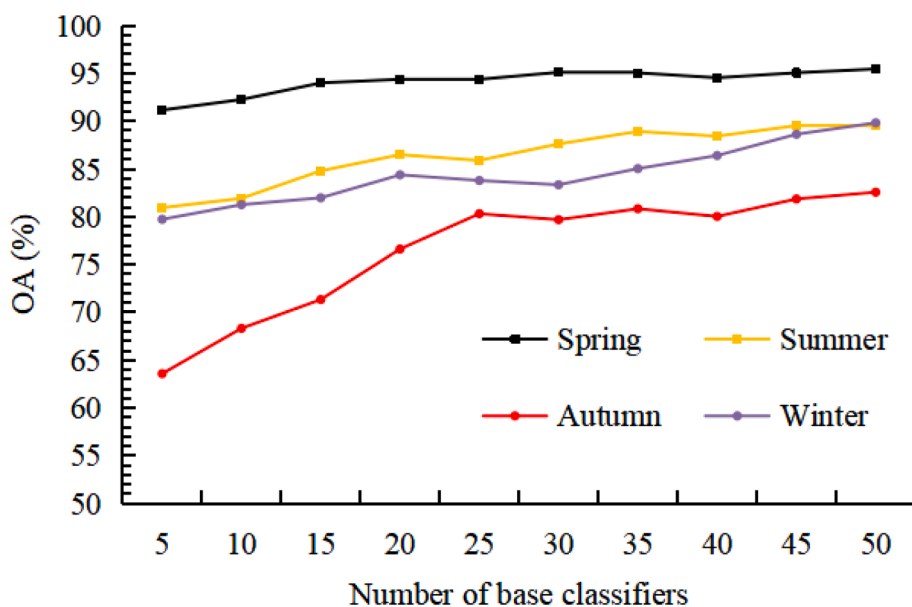**Fig. 11.** Effect of training sample size on classification accuracy.

Fig. 12. Effect of the number of shallow classifiers on the OA of CART-based FRM_DSL.

are influenced by factors such as samples, image acquisition time, and classification methods. While our product boasts the highest overall accuracy, it encounters challenges in accurately classifying GR due to the potential confusion between GR and SH, as well as CL. The fusion of classification results from four quarterly periods in our product introduces a certain level of ambiguity, leading to some errors in

Table 4
Classification accuracy of RF-based FRM_DSL with different number of CARTs (%).

Dataset	Number of CARTs in RF				
	10	20	30	40	50
Spring	95.36	95.69	95.99	96.03	96.55
Summer	89.60	89.98	90.27	90.66	90.95
Autumn	81.65	82.99	83.56	84.33	84.66
Winter	88.93	89.79	90.22	90.68	91.22

classifying unstable LULC types. However, for more stable LULC types like CL, FO, WB, and IS, our product still achieves a commendable classification performance.

6. Conclusions

In the process of large-scale remote sensing image classification, the extraction of stable features plays a crucial role in constructing a classifier with strong generalization ability. In this study, we have discovered that generating EFM for each pixel, coupled with deep learning utilizing CNN, not only preserves detailed ground object information but also yields more stable features. Additionally, continuing deep learning based on the classification results obtained from shallow learning using MCS allows for the extraction of relatively stable features of ground objects. FRM_DSL integrates these two approaches to obtain a deep learning classifier with enhanced generalization ability. It successfully classifies Landsat OLI images across a large extent in the study region, namely the YREB, and achieves high accuracy in extracting LULC information.

A meticulous comparative analysis of various classifiers reveals that FRM_DSL exhibits a remarkable ability to achieve high classification accuracy with a relatively smaller sample size. As the number of samples increases, the accuracy of FRM_DSL becomes even more pronounced. Additionally, the performance of FRM_DSL is greatly influenced by the number and accuracy of its base classifiers in the MCS module. When a larger number of base classifiers are utilized, FRM_DSL demonstrates higher accuracy. Furthermore, by incorporating more accurate base classifiers, the classification accuracy of FRM_DSL can be further
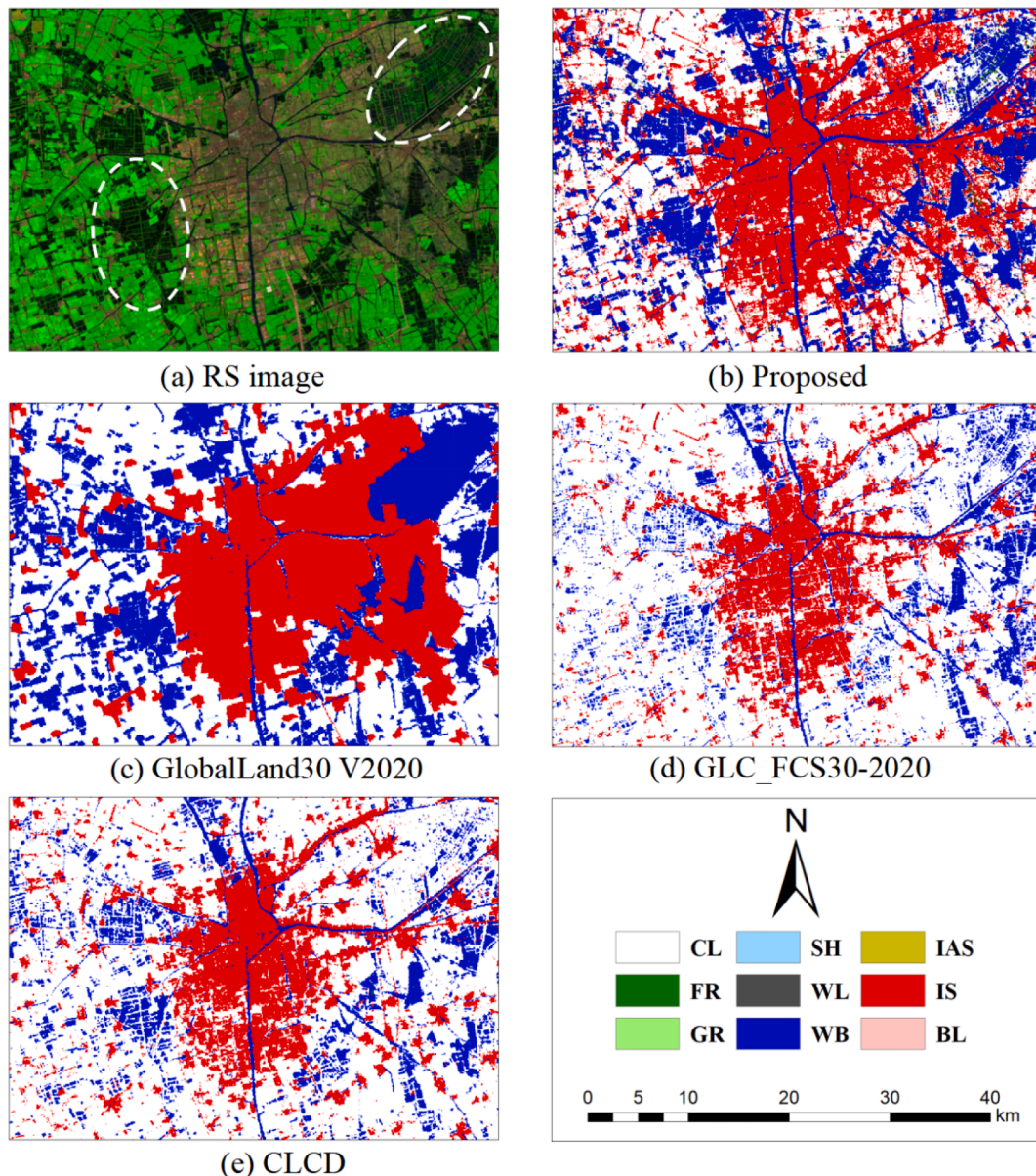

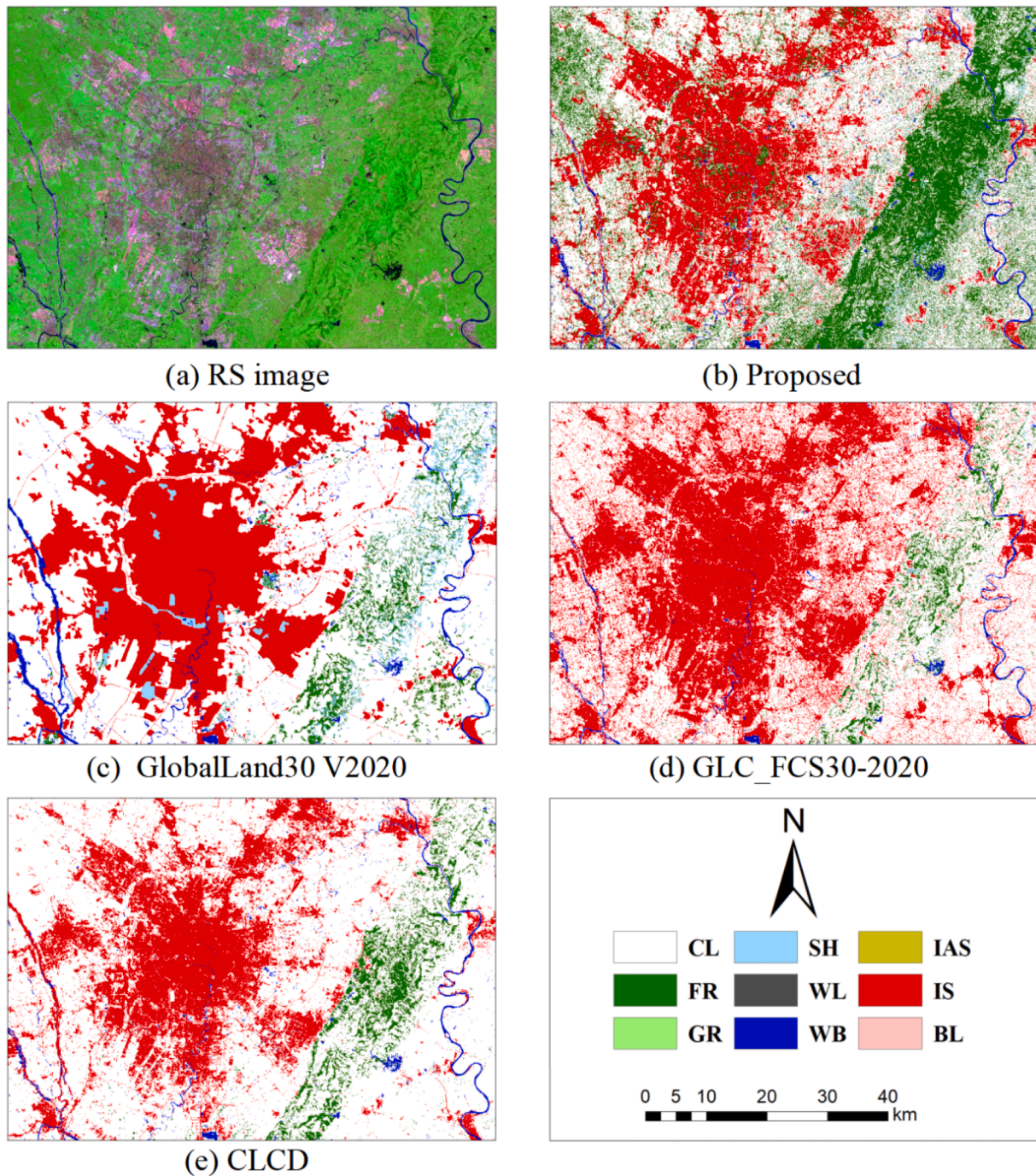**Fig. 13.** LULC mapping of Scene 1.

Fig. 14. LULC mapping of Scene 2.

enhanced.

In comparison to existing contemporaneous LULC products in 2020, FRM_DSL demonstrates a commendable performance in LULC mapping. As a result, the trained model in this study holds potential for LULC extraction using Landsat 8 OLI images in different regions or time periods. Furthermore, with the advent of Landsat 9 OLI-2, which is now publicly available, utilizing the currently trained model to classify OLI-2 data and obtain the latest LULC products presents a promising direction for future research continuation.

CRediT authorship contribution statement

Peng Dou: Writing – original draft, Project administration, Methodology, Funding acquisition, Conceptualization. **Huanfeng Shen:** Writing – review & editing, Conceptualization. **Chunlin Huang:** Validation, Supervision. **Zhiwei Li:** Software, Resources, Investigation. **Yujun Mao:** Visualization, Data curation. **Xinghua Li:** Formal analysis.
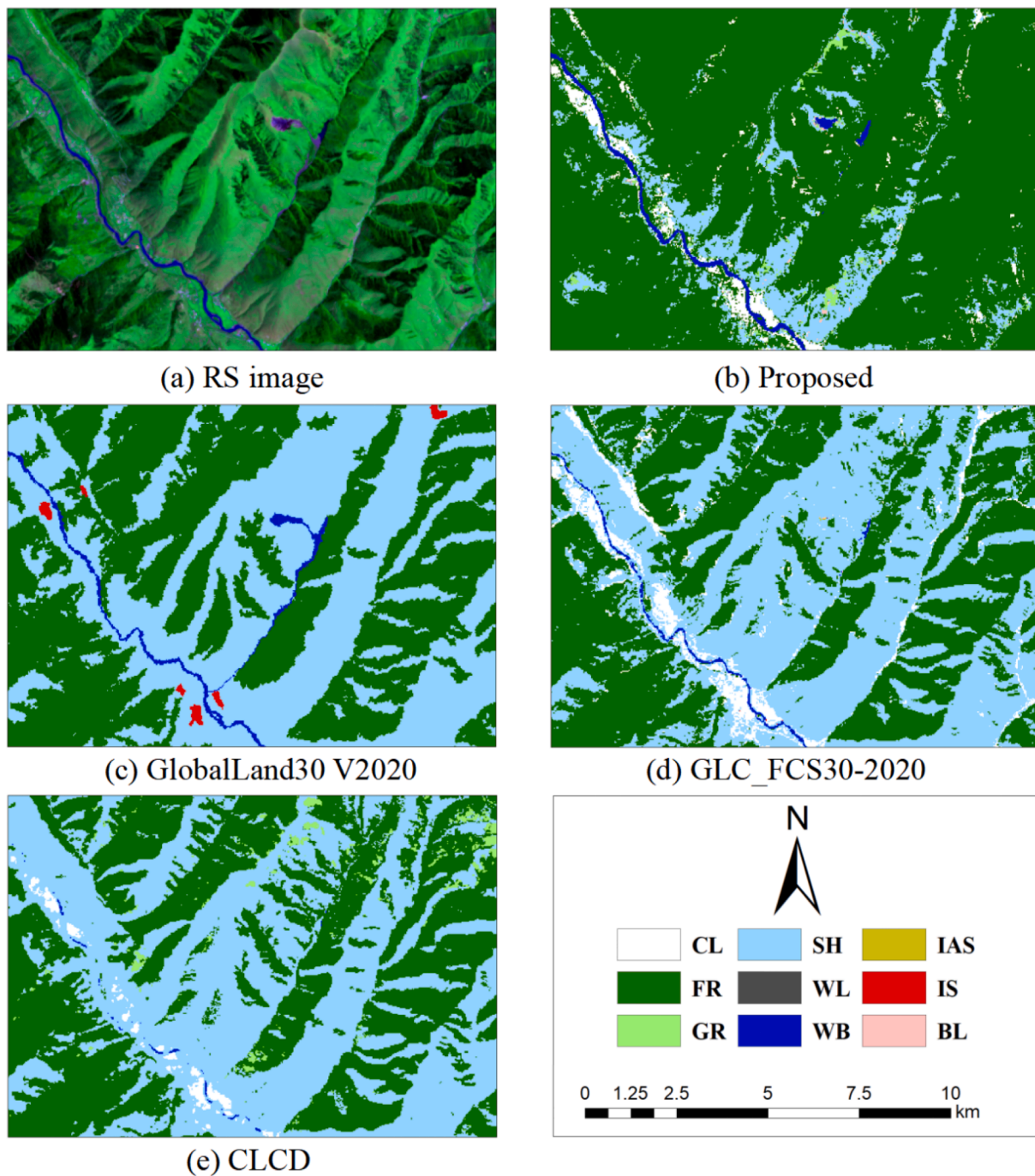
Fig. 15. LULC mapping of Scene 3.

Table 5
Evaluations of the accuracy for different LULC datasets (%).

LULC		CL	FO	GR	SH	WL	WB	IS	BL	LAS	OA	Kappa
Proposed	PA	96.30	93.32	16.80	45.10	28.88	91.98	85.19	44.68	91.17	84.77	0.786
	UA	81.70	97.58	28.83	42.37	59.58	75.71	87.04	68.28	50.05		
CLCD	PA	76.13	95.10	42.09	49.60	25.60	85.83	86.45	45.42	80.31	83.49	0.762
	UA	95.27	94.94	56.13	48.03	45.60	63.43	71.00	17.07	21.59		
GLC_FCS30-2020	PA	73.89	94.08	25.63	40.42	18.98	85.85	85.57	54.75	84.71	80.03	0.715
	UA	92.15	90.08	45.35	47.01	8.83	54.07	78.16	2.95	25.41		
GlobalLand30 V2020	PA	72.88	92.20	20.52	45.39	59.77	75.29	80.15	46.69	92.72	76.89	0.675
	UA	87.13	84.31	55.25	54.34	16.78	65.81	68.83	9.45	23.06		

15

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

Appendix. Table of full names and abbreviations covered in this paper

Full name	Abbreviation
Convolutional neural network	CNN
Land Use/Land Cover	LULC
Extent feature map	EFM
Multiple classifiers system	MCS
Yangtze River Economic Belt	YREB
Remote sensing	RS
Deep-shallow learning	DSL
Feature relationships matrix-based deep-shallow learning	FRM_DSL
Feature relationship matrix	FRM
Residual network	ResNet
Regions of interest	ROIs
Space-spectrum joint classification based on feature relationship matrix	SSJC_ERM
Classification and Regression Trees	CART
Lizhiwei Net	LZWNet
Spectral-Spatial Fully Convolutional Networks	SSFCN
One-dimensional CNN	1D CNN
Random forest	RF
Generative Extreme Learning Machine	GenELM
Batch size	BS
Overall accuracy	OA
Local accuracy	LA
Non-local accuracy	NLA
Cropland	CL
Forest	FO
Grassland	GR
Shrub	SH
Wetland	WL
Water body	WB
Impervious surface	IS
Bare land	BL
Ice and snow	IAS
China Land Cover Dataset	CLCD
Global Land Cover with Fine Classification System at 30 m in 2020	GLC_FCS30-2020

References

Abdali, E., Valadan Zoej, M.J., Taheri Dehkordi, A., Ghaderpour, E., 2024. A parallel-cascaded ensemble of machine learning models for crop type classification in Google Earth Engine using multi-temporal Sentinel-1/2 and Landsat-8/9 remote sensing data. Remote Sens. (Basel) 16, 127. https://doi.org/10.3390/rs16010127.

Breiman, L., 2001. Random Forests. Mach. Learn. 45, 5–32. https://doi.org/10.1023/A:1010933404324.

Brovelli, M., Molinari, M., Hussein, E., Chen, J., Li, R., 2015. The First Comprehensive Accuracy Assessment of GlobeLand30 at a National Level: Methodology and Results. Remote Sens. (Basel) 7, 4191–4212. https://doi.org/10.3390/rs70404191.

Cheng, X., He, X., Qiao, M., Li, P., Hu, S., Chang, P., Tian, Z., 2022. Enhanced contextual representation with deep neural networks for land cover classification based on remote sensing images. Int. J. Appl. Earth Obs. Geoinf. 107, 102706.

Cui, W., Zheng, Z., Zhou, Q., Huang, J., Yuan, Y., 2018. Application of a parallel spectral–spatial convolution neural network in object-oriented remote sensing land use classification. Remote Sens. Lett. 9 (4), 334–342.

Davydzenka, T., Tahmasebi, P., Carroll, M., 2022. Improving remote sensing classification: a deep-learning-assisted model. Comput. Geosci. 164, 105123 https://doi.org/10.1016/j.cageo.2022.105123.

DeLancey, E.R., Simms, J.F., Mahdianpari, M., Brisco, B., Mahoney, C., Kariyeva, J., 2019. Comparing deep learning and shallow learning for large-scale wetland classification in Alberta, Canada. Remote Sens. 12, 2. https://doi.org/10.3390/rs12010002.

Dou, P., Shen, H., Li, Z., Guan, X., Huang, W., 2021a. Remote sensing image classification using deep-shallow learning. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 14, 3070–3083. https://doi.org/10.1109/JSTARS.2021.3062635.

Dou, P., Shen, H., Li, Z., Guan, X., 2021b. Time series remote sensing image classification framework using combination of deep learning and multiple classifiers system. Int. J. Appl. Earth Obs. Geoinf. https://doi.org/10.1016/j.jag.2021.102477.

Dou, P., Huang, C., Han, W., Hou, J., Zhang, Y., Gu, J., 2024. Remote sensing image classification using an ensemble framework without multiple classifiers. ISPRS J. Photogramm. Remote Sens. 208, 190–209. https://doi.org/10.1016/j.isprsjprs.2023.12.012.

Dou, P., Zeng, C., 2020. Hyperspectral image classification using feature relations map learning. Remote Sens. (Basel) 12, 2956. https://doi.org/10.3390/rs12182956.

Du, P., Bai, X., Tan, K., Xue, Z., Samat, A., Xia, J., Li, E., Su, H., Liu, W., 2020. Advances of four machine learning methods for spatial data handling: a review. J. Geovis. Spatial Anal. 4 (1) https://doi.org/10.1007/s41651-020-00048-5.

Gao, H., Yao, D., Yang, Y., Li, C., Liu, H., Hua, Z., 2020. Multiscale 3-D-CNN based on spatial–spectral joint feature extraction for hyperspectral remote sensing images classification. J. Electron. Imaging 29, 1. https://doi.org/10.1117/1.JEI.29.1.013007.

Gopal Singh, P., Bordu, N., Singh, D., Yahia, H., Daoudi, K., 2020. Permuted spectral and permuted spectral-spatial CNN models for PolSAR-multispectral data based land cover classification. Int. J. Remote Sens. 42, 1096–1120. https://doi.org/10.1080/01431161.2020.1823041.

Han, D., Huang, G., Liu, L., Zhai, M., Gao, S., 2021. Multi-regional industrial wastewater metabolism analysis for the Yangtze River Economic Belt, China. Environ. Pollut. 284, 117118 https://doi.org/10.1016/j.envpol.2021.117118.

Homer, C., Dewitz, J., Jin, S., Xian, G., Costello, C., Danielson, P., Gass, L., Funk, M., Wickham, J., Stehman, S., Auch, R., Riitters, K., 2020. Conterminous United States land cover change patterns 2001–2016 from the 2016 National Land Cover Database. ISPRS J. Photogramm. Remote Sens. https://doi.org/10.1016/j.isprsjprs.2020.02.019.

Li, F., Liu, X., Liao, S., Jia, P., 2021. The modified normalized urban area composite index: a satellite-derived high-resolution index for extracting urban areas. Remote Sens. (Basel) 13, 2350. https://doi.org/10.3390/rs13122350.

Li, X., Wang, L., Guan, H., et al., 2024. Urban tree species classification using UAV-based multispectral images and LiDAR point clouds. J. Geovis. Spatial Anal 8, 5. https://doi.org/10.1007/s41651-023-00167-9.

Maggiolo, L., Marcos, D., Moser, G., Serpico, S.B., Tuia, D., 2022. A semisupervised CRF model for CNN-based semantic segmentation with sparse ground truth. IEEE Trans. Geosci. Remote Sens. 60, 1–15. https://doi.org/10.1109/TGRS.2021.3095832.

Mohith, M., Karthi, R., 2022. Land use and land cover mapping of Landsat image using segmentation techniques. Lect. Notes Electr. Eng. 717–728. https://doi.org/10.1007/978-981-19-2177-3_67.

Pan, E., Ma, Y., Mei, X., et al., 2019. Spectral-spatial classification of hyperspectral image based on a joint attention network. In: IGARSS 2019–2019 IEEE International Geoscience and Remote Sensing Symposium. https://doi.org/10.1109/IGARSS.2019.8898758.

Paoletti, M.E., Haut, J.M., Plaza, J., Plaza, A., 2018. A new deep convolutional neural network for fast hyperspectral image classification. ISPRS J. Photogramm. Remote Sens. 145, 120–147. https://doi.org/10.1016/j.isprsjprs.2017.11.021.

Pasquarella, V.J., Holden, C.E., Woodcock, C.E., 2018. Improved mapping of forest type using spectral-temporal Landsat features. Remote Sens. Environ. 210, 193–207. https://doi.org/10.1016/j.rse.2018.02.064.

Phiri, D., Morgenroth, J., Xu, C., Hermosilla, T., 2018. Effects of pre-processing methods on Landsat OLI-8 land cover classification using OBIA and random forests classifier. Int. J. Appl. Earth Obs. Geoinf. 73, 170–178.

Sharma, A., Liu, X., Yang, X., 2018. Land cover classification from multi-temporal, multi-spectral remotely sensed imagery using patch-based recurrent neural networks. Neural Netw. 105, 346–355. https://doi.org/10.1016/j.neunet.2018.05.019.

Talukdar, S., Singha, P., Mahato, S., Shahfahad, P., Liou, Y.-A., Rahman, A., 2020. Land-use land-cover classification by machine learning classifiers for satellite observations—a review. Remote Sens. 2020 (12), 1135 https://doi.org/10.3390/rs12071135.

Tieleman, T., Hinton, G., 2012. Lecture 6.5-rmsprop: divide the gradient by a running average of its recent magnitude. COURSERA: Neural Netw. Mach. Learn. 4 (2), 26–31.

Wang, Y., Li, Z., Zeng, C., Xia, G.-S., Shen, H., 2020. An urban water extraction method combining deep learning and Google Earth Engine. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 13, 769–782. https://doi.org/10.1109/JSTARS.2020.2971783.

Xu, Y., Du, B., Zhang, L., 2020. Beyond the patchwise classification: spectral-spatial fully convolutional networks for hyperspectral image classification. IEEE Trans. Big Data. https://doi.org/10.1109/tbdata.2019.2923243.

Xu, Z., Guan, K., Casler, N., Peng, B., Wang, S., 2018b. A 3D convolutional neural network method for land cover classification using LiDAR and multi-temporal Landsat imagery. ISPRS J. Photogramm. Remote Sens. 144, 423–434. https://doi.org/10.1016/j.isprsjprs.2018.08.005.

Xu, K., Huang, H., Deng, P., 2022. Remote sensing image scene classification based on global-local dual-branch structure model. IEEE Geosci. Remote Sens. Lett. 19, 1–5. https://doi.org/10.1109/LGRS.2021.3075712.

Xu, X., Yang, G., Tan, Y., Liu, J., Hu, H., 2018a. Ecosystem services trade-offs and determinants in China's Yangtze River Economic Belt from 2000 to 2015. Sci. Total Environ. 634, 1601–1614. https://doi.org/10.1016/j.scitotenv.2018.04.046.

Yang, J., Huang, X., 2021. The 30 m annual land cover dataset and its dynamics in China from 1990 to 2019. Earth Syst. Sci. Data 13, 3907–3925. https://doi.org/10.5194/essd-13-3907-2021.

Yu, L., Wang, J., Gong, P., 2013. Improving 30 m global land-cover map FROM-GLC with time series MODIS and auxiliary data sets: a segmentation-based approach. Int. J. Remote Sens. 34, 5851–5867. https://doi.org/10.1080/01431161.2013.798055.

Zhang, M., Lin, H., Wang, G., Sun, H., Fu, J., 2018. Mapping paddy rice using a convolutional neural network (CNN) with Landsat 8 datasets in the Dongting Lake Area, China. Remote Sens. 10, 1840 https://doi.org/10.3390/rs10111840.

Zhang, E., Liu, L., Huang, L., Ng, K.S., 2021a. An automated, generalized, deep-learning-based method for delineating the calving fronts of Greenland glaciers from multi-sensor remote sensing imagery. Remote Sens. Environ. 254, 112265 https://doi.org/10.1016/j.rse.2020.112265.

Zhang, X., Liu, L., Chen, X., Gao, Y., Xie, S., Mi, J., 2021b. GLC_FCS30: global land-cover product with fine classification system at 30 m using time-series Landsat imagery. Earth Syst. Sci. Data 13, 2753–2776. https://doi.org/10.5194/essd-13-2753-2021.

Zhang, Z., Liu, D., Gao, D., Shi, G., 2022b. S^{3}Net: spectral–spatial–semantic network for hyperspectral image classification with the multiway attention mechanism. IEEE Trans. Geosci. Remote Sens. 60, 1–17. https://doi.org/10.1109/TGRS.2021.3067356.

Zhang, F., Yan, M., Hu, C., Ni, J., Zhou, Y., 2022a. Integrating coordinate features in CNN-based remote sensing imagery classification. IEEE Geosci. Remote Sens. Lett. 19, 1–5. https://doi.org/10.1109/LGRS.2020.3045744.

Zhao, X., Gao, L., Chen, Z., Zhang, B., Liao, W., Yang, X., 2019. An entropy and MRF model-based CNN for large-scale Landsat image classification. IEEE Geosci. Remote Sens. Lett. 16, 1145–1149. https://doi.org/10.1109/LGRS.2019.2890996.

Zhong, Z., Li, J., Luo, Z., Chapman, M., 2018. Spectral-spatial residual network for hyperspectral image classification: a 3-D deep learning framework. IEEE Trans. Geosci. Remote Sens. 56, 847–858. https://doi.org/10.1109/TGRS.2017.2755542.