International Journal of Applied Earth Observation and Geoinformation 129 (2024) 103780

Contents lists available at ScienceDirect

International Journal of Applied Earth Observation and Geoinformation

journal homepage: www.elsevier.com/locate/jag

Fusion of GaoFen-5 and Sentinel-2B data for lithological mapping using vision transformer dynamic graph convolutional network

Yanni Dong [a], Zhenzhen Yang [b], Quanwei Liu [c], Renguang Zuo [d], Ziye Wang [d,*]

[a] *School of Resource and Environmental Sciences, Wuhan University, Wuhan 430079, China*
[b] *School of Geophysics and Geomatics, China University of Geosciences, Wuhan 430074, China*
[c] *College of Science and Engineering, James Cook University, Cairns QLD 4878, Australia*
[d] *State Key Laboratory of Geological Processes and Mineral Resources, China University of Geosciences, Wuhan 430074, China*

ARTICLE INFO

Keywords:
Lithological mapping
Data fusion
Vision transformer
Graph convolutional network

ABSTRACT

Lithological identification and mapping using remote sensing (RS) imagery are challenging. Traditional lithological mapping relies mainly on multispectral data and machine learning methods. However, inadequate spectral information and inappropriate classification algorithms are major problems for RS geological applications. Moreover, satellite hyperspectral images (HSI) at low spatial resolution and convolutional neural network (CNN)-based methods with incomplete feature extraction remain challenging because of the limitations of sensor imaging and convolutional kernels for lithological mapping. To address the above issues, in this study, smoothing filter-based intensity modulation (SFIM) fusion technology is first employed to fuse GaoFen-5 hyperspectral images and Sentinel-2B multispectral images. This approach significantly improves spatial details and enriches spectral information. Subsequently, a novel Vision Transformer Dynamic Graph Convolutional Network (ViT-DGCN) is proposed for lithological mapping of the Cuonadong dome, Tibet, China. ViT-DGCN is a joint model consisting of a transformer and a dynamic graph convolution module that enhances feature extraction capabilities by exploring long-range interaction sequence features and dynamic graph structure information in a targeted manner. The proposed algorithm exhibits superior performance compared to the others, achieving an overall accuracy of 97% for the Cuonadong dome using only 1% of the available training samples.

1. Introduction

As a basic task in regional geological surveys, lithological mapping can be used to determine the relationship between the formation and storage of minerals and the lithology (Lyon, 1972), which has important guiding significance in the study of regional geological minerals (Fan and Wang, 2020; Wan et al., 2021). However, the lithological mapping of areas with complex geological environments and inaccessible areas requires considerable manpower, materials, and financial resources. Lithological mapping based on remote sensing (RS) technology is an accessible and low-cost solution that provides an efficient means of identifying lithologies (Girija and Mayappan, 2019).

Traditional RS lithological mapping relies mostly on multispectral images (MSI), such as Landsat satellites data, WorldView-3 data, Sentinel-2 multispectral data, and advanced space-borne thermal emission and reflection radiometer (ASTER) data (Ye et al., 2017; Ge et al., 2018; Xi et al., 2022; Ousmanou et al., 2023). Although MSI data perform well in lithological classification, insufficient spectral information of the MSI data is the main limitation. As a current frontier field of RS technology, hyperspectral technology uses hundreds of narrow bands for the continuous imaging of ground objects to obtain rich spectral information, which is more conducive to identifying and classifying mineral and rock units (Peighambari and Zhang, 2021). Thus, lithological mapping based on hyperspectral images (HSI) has received attention in recent years to address the low spectral resolution of MSI data. There are some precedents for utilizing HSI to update lithological maps, exemplified by the application of data from Hyperion (Zhang and Li, 2014), GaoFen-5 (GF-5) (Ye et al., 2020), and ZY-1 02D satellites (Yu et al., 2021), These instances are not only noteworthy but also suggest important and meaningful directions for future applications. However, increases in spectral resolution often result in reduced spatial resolution due to sensor imaging limitations. All the aforementioned satellite-based data exhibit a spatial resolution of 30 m. Thus, satellite-based high-resolution HSI data are rarely accessible, and the inadequate spatial

* Corresponding author.
E-mail address: ziyewang@cug.edu.cn (Z. Wang).

https://doi.org/10.1016/j.jag.2024.103780
Received 24 October 2023; Received in revised form 10 March 2024; Accepted 17 March 2024
Available online 28 March 2024

resolution of HSI compromises the advantages of practical application (Tong et al., 2016).

Technologically, for a single type of RS data, an increase in the spectral resolution leads to a relative decrease in the spatial resolution, resulting in greater image element mixing, which affects the mineral identification effect (Liu et al., 2018). Thus, image fusion techniques for multiple RS data provide an effective solution for regional lithological mapping (Yang et al., 2018; Manap and Bekir, 2022; Khashaba et al., 2023). A typical example is based on HSI and MSI fusion techniques, in which MSI data often have a higher spatial resolution than HSI data. Fusing low spatial resolution HSI and high spatial resolution MSI can yield new data with both high spatial and spectral resolution, which can provide more accurate geological mapping applications than all available individual data (Vivone, 2023).

The most common HSI and MSI image fusion techniques are the component substitution methods. Component substitution methods enhance spatial resolution via the projection transformation. Such as Gram-Schmidt (GS), intensity-hue-saturation (IHS), Brovey transform (BT), and decimated wavelet transform (DWT) (Yokoya et al., 2017; Zhang et al., 2023). Nevertheless, these methods exemplified by GS often suffer from spectral distortion due to the component substitution changing the spectral properties of the original data (Sun et al., 2022a). Thus, difficulties remain in HSI and MSI image fusion for lithological classification owing to serious spectral distortion and noise (Pal et al., 2020). A common solution is to inject spatial details from a high-resolution image into a low-resolution image depending on their correlation (Ren et al., 2020). Inspired by this, this study applied a smoothing filter-based intensity modulation (SFIM) for spectral preservation fusion technique (Liu, 2000) to fuse the high spatial details of MSI (i.e., Sentinel-2B data) with the enriched spectral information of HSI (i.e., GF-5 data). The SFIM method minimizes spectral distortions and maximizes the retention of spectral information while improving the spatial quality of HSI data (Kabolizadeh et al., 2022).

A series of lithological mapping algorithms were proposed based on the acquired RS images. Machine learning (ML)-based approaches including metric learning, support vector machine (SVM) and random forest (RF) have been successfully used in lithological mapping. (Bachri et al., 2019; Girija and Mayappan, 2019; Wang et al., 2020; Latif et al., 2023). Although these studies have shown that this lithological mapping concept has particular advantages under specific geological conditions, they only focused on shallow spectral features, ignoring the spatial relationship among pixels.

With the continuous advancement and expansion of deep learning (DL) techniques, recent studies have been conducted on deep feature extraction for lithological mapping (Shirmard et al., 2022b; Kim et al., 2022). Examples include autoencoder (AE) (Xiong et al., 2022), generative adversarial networks (Zhang et al., 2021), convolutional neural networks (CNN) (Brandmeier and Chen, 2019; Wang et al., 2021a; Pan et al., 2023), graph convolutional networks (GCN) (Zuo and Xu, 2023). These DL methods show considerable potential for extracting spatial information from geological big data to support lithological mapping (Wang and Zuo, 2022; Wang et al., 2023a). However, most existing CNN-based lithological mapping methods suffer from the locality limitation of convolutional kernel operations, which cannot fully extract the relations of hidden rock bodies and ignore global information (Hong et al., 2021; Zou et al., 2022; Dang et al., 2023).

To promote better spatial-spectral features, a stronger Vision Transformer (ViT) model has been successfully applied from natural language processing (NLP) in the RS field. ViT can capture high-level feature representations of input data, and shows superior performance in HSI classification (Qing et al., 2021; Sun et al., 2022b; Dang et al., 2023). The key module of the transformer is called the multi-head self-attention (MHSA) mechanism, which has the capability of modeling content-dependent long-distance interactions to efficiently extract sequence features (Yang et al., 2022). Nevertheless, ViT does not generalize well for training with an insufficient data volume (Liu et al., 2021; Lee et al., 2023). To alleviate this problem, we consider HSI classification from a graphical perspective, because GCN can construct graph-structured features and establish connections between them in non-Euclidean space (Kipf and Welling, 2016). GCN can effectively reduce the large amount of training data required for the transformer and achieve good classification results (Liu et al., 2022).

This study provides an alternative approach for lithological mapping in the Cuonadong dome study area using HS and MS data fusion technology and a lithology identification model based on ViT. A large Sn-Be-W deposit at the Cuonadong dome has been discovered and reported (Li et al., 2017). First, we integrate Sentinel-2B and GF-5 RS images by applying the SFIM fusion algorithm, producing better-quality data with both high spatial and spectral resolutions. Second, the fused data are entered into the designed vision transformer dynamic graph convolutional network (ViT-DGCN) model to classify seven lithological categories and map lithological results of the Cuonadong dome. ViT is used for the first time in HSI lithological mapping and has never been performed in this field. On this basis, we design a dynamic GCN module and integrate it into ViT, which can adaptively aggregate features with a graph view, thereby accelerating the feature extraction efficiency. The ViT-DGCN model can obtain a highly accurate lithological classification task using a small amount of training data. This work's main contributions include:

1. To further enhance the spatial information of the HSI data, we use the SFIM algorithm to fuse Sentinel-2B with GF-5 data for a more accurate description of spatial information. The classification results demonstrate the excellent performance of the fusion process.
2. To address the problem of inadequate feature extraction in CNN for lithological mapping, we design a ViT-DGCN model with transformer and dynamic GCN module that can fully learn sequence features and adaptively extract topological characteristics of input data.
3. The proposed ViT-DGCN model is illustrated for lithological mapping in the Cuonadong dome, Tibet, China, using GF-5 and Sentinel-2B fusion imagery. Satisfactory results are obtained compared with other state-of-the-art DL methods.

The remainder of this paper is organized as follows the following structure. Section 2 presents a detailed literature review of previous work. Section 3 presents a geological overview of the study area and the data used. Section 4 focuses on the data preprocessing, data fusion processes, and the proposed ViT-DGCN. Section 5 describes the experimental detail. Section 6 analyses the lithological classification results of the comparison methods. The performance of the model in terms of the sample size and generalization power is discussed in Section 7. Finally, Section 8 concludes the paper and presents the prospects for future research.

2. Literature review

Lithological mapping is essential for determining the targeted lithological units within surrounding rocks, and is one of the fundamental approaches to the geological evaluation concerning of mineral exploration (Shirmard et al., 2022a). RS techniques and digital satellite image processing are now increasingly used to prepare geological and lithological maps of land surfaces (Bhan and Krishnanunni, 1983; Bachri et al., 2019; Shayeganpour et al., 2021; Lee et al., 2024). In addition, low-cost RS methods are used to map areas that are difficult to access and provide an excellent alternative to routine fieldwork. Based on the principles of RS, diverse materials reflect unique electromagnetic energy owing to their physicochemical features, which help to identify the spectral properties of rock minerals that are necessary to automate lithological classification (Bachri et al., 2019). In this section, we review some key developments regarding datasets and methods similar to those used.

2.1. Lithological mapping using multiple RS data

In the RS dataset, the most commonly used were MSI data, including Landsat-8 OLI, ASTER, and Sentinel-2 (Ge et al., 2018), which effectively show the lithological mapping of rock units that were previously difficult to reveal using optical images. These studies demonstrated the successful application of RS lithological mapping (El-Omairi and El Garouani, 2023), however, there were some limitations and shortcomings of the single RS data. The integration of various sources of RS data is conducive to improving the accuracy of lithological maps and provides more useful and complete information (Manap and Bekir, 2022). For example, Bachri et al (2019) combined Landsat 8 and digital elevation model (DEM) data for automatic lithological mapping. Kabolizadeh et al. (2022) applied an optimum fusion method using sentinel-2 and ASTER images to improve the lithological mapping of sedimentary rocks. Marzouki and Dridri (2023) used Landsat 8 and ASTER data for lithological discrimination in a Tiwit case study.

Despite its wide utilization, MSI has limitations owing to its lower spectral resolution, whereas HSI opens new possibilities for lithological mapping (Ding and Ding, 2022). There are many precedents for using HSI to map lithological classes. For example, Zhang and Li (2014) improved the spectral angle mapper for lithological mapping using Earth Observing-1 Hyperion data with a spatial resolution of 30 m. Ye et al. (2020) applied different convolutional CNN methods and explored the potential and applicability based on GF-5 hyperspectral data of 30 m Yu et al. (2021) processed ZY-1 02D data (30 m) for lithological mapping in the Liuyuan area and achieved an excellent performance. Nevertheless, the lower spatial resolution of HSI data is a minor disadvantage compared with MSI data. Inspired by the idea of data fusion, we integrated MSI and HSI for lithological mapping. MSI data behaves better in the fusion of HSI data, because of higher correlations of their spectral bands (Sun et al., 2019). Considering the advantages of the higher spectral resolution of GF-5 data and higher spatial resolution of Sentinel-2B (S2B) data, this study uses S2B and GF-5 images to enhance lithological mapping.

2.2. Lithological mapping based on DL methods

In the methods, due to the popularity of DL research (Wang et al., 2022a; Deng et al., 2023), many approaches for the lithological mapping task have adopted DNNs currently. These networks were used to extract high-level features, which were then used to map various lithological categories. Recent methodological advancements in geoscience mainly focused on using CNN algorithms for lithological mapping tasks (Shirmard et al., 2022b). Ye et al. (2020) explored the potential and applicability of CNN for lithological mapping. This was achieved by evaluating the classification performance of the multi-scale 3D deep CNN, hybrid spectral CNN, and spectral–spatial unified network on the GF-5 data. Wang et al. (2021a) integrated multi-source geological data, including ASTER images, DEM, geochemical, and aeromagnetic data, along with a fully CNN for lithological mapping. Their findings demonstrated the efficacy of incorporating multi-source data and CNN in identifying lithological features. Yu et al. (2021) introduced an unsupervised 3D convolutional autoencoder algorithm to process ZY-1 02D data for lithological mapping of the Liuyuan area. The algorithm achieved outstanding performance in generating lithological maps. Wang et al. (2023a) proposed an adversarial semi-supervised segmentation network based on RS data to obtain a lithological map. CNN has proven to be a powerful feature extractor for lithological classification (Wang et al., 2023b). However, because of the limitations of the inherent network structure and convolution kernel operation, CNN cannot adequately explore and represent the sequential attributes of spectral features. CNN focuses overly on local spatial content information, which may neglect global sequential information in the learned features at the spectral level.

The development of the transformer model brings a new approach to image classification. Dosovitskiy et al. (2020) proposed the first transformer-based model for images, named ViT, which demonstrated a strong performance. The transformer is very effective at processing sequence data and can extract global features of the input data through the MHSA. However, the simple application of the ViT model to HSI classification has many limitations (Dang et al., 2023). The ViT model achieved excellent results only when it relied on a large amount of training sample data. To address this limitation, we introduced a dynamic GCN (Liu et al., 2022), as it can mitigate the model's reliance on the number of training samples. Unlike classic CNN, we propose a modified transformer model with a dynamic GCN that focuses on the channel relationships and topological features. This transformer-based model was first applied to lithological mapping, which is a novel exploration of the geological applications of transformer.

3. Study area and data description

The Cuonadong dome is located approximately 20 km north of the Tibet South Demolition System in the Himalayan orogenic belt of the Tibet Autonomous Region, in the northern part of Cuona County, Shannan City, as shown in Fig. 1. The dome covers approximately 600 km^2 (Wang et al., 2020). Research on Himalayan leucogranites has always been a hot topic in geosciences, especially in rare metal mineralization (Cao et al., 2022). With further development of prospecting and exploration, geologists have discovered that the Cuonadong dome is the first rare Sn-Wu-Be polymetallic deposit in the Himalayan region to be surveyed and circled for rare metal ore bodies (Cao et al., 2021; Wang et al., 2022b). The predicted reserves of BeO exceed 500,000 tons, manifesting as potential exploration targets for rare metal deposits in Himalayan leucogranite (Li et al., 2017). Since the Cuonadong deposit has a mineralization background similar to that of a large number of mineralized areas in the Himalayas, it is important to consider Cuonadong as a research area. Thus, it is meaningful but challenging to map the distribution of lithology in the study area.

Seven lithological units have been detected in the Cuonadong area: Jurassic sandstone and slate, Quaternary strata, Early Paleozoic marble, Triassic sandstone and slate, Cambrian granitic gneiss, Paleozoic biotite quartz schist, and Himalayan leucogranite (Wang et al., 2021b). The typical lithological features and major ore minerals of the Cuonadong dome are shown in Table 1. Fig. 2 shows the spectral profiles of several rocks in the visible and near-infrared regions. The spectral characteristics of rock minerals are related to their mineral composition, weathering characteristics, and structural backgrounds. These petrographic properties determine the special spectral characteristics of rocks or minerals and create a theoretical basis for the identification of rock bodies based on RS images.

Two types of RS data were obtained for lithological mapping S2B and GF-5, as shown in Table 2. S2B covers 13 bands with varying resolutions from the visible and near-infrared (VNIR) to the short-wave infrared (SWIR) ranges, providing a maximum spatial resolution of 10 m. Sentinel-2 data are available for free from the European Space Agency (https://dataspace.copernicus.eu/). The GF-5 HSI includes 330 bands ranging from the VNIR to SWIR and has a spatial resolution of 30 m (Liu et al., 2019) (https://geogf.agrs.cn/search/). S2B data were acquired on December 02, 2021, and GF-5 data were acquired on October 13, 2019.

4. Methodology

In this section, we begin by illustrating the RS data used for lithological mapping and their pre-processing steps. Then, we perform the data fusion process, after which, the fused data are fed into the ViT-DGCN model and finally output the lithological classification results. Fig. 3 illustrates the framework of this study.

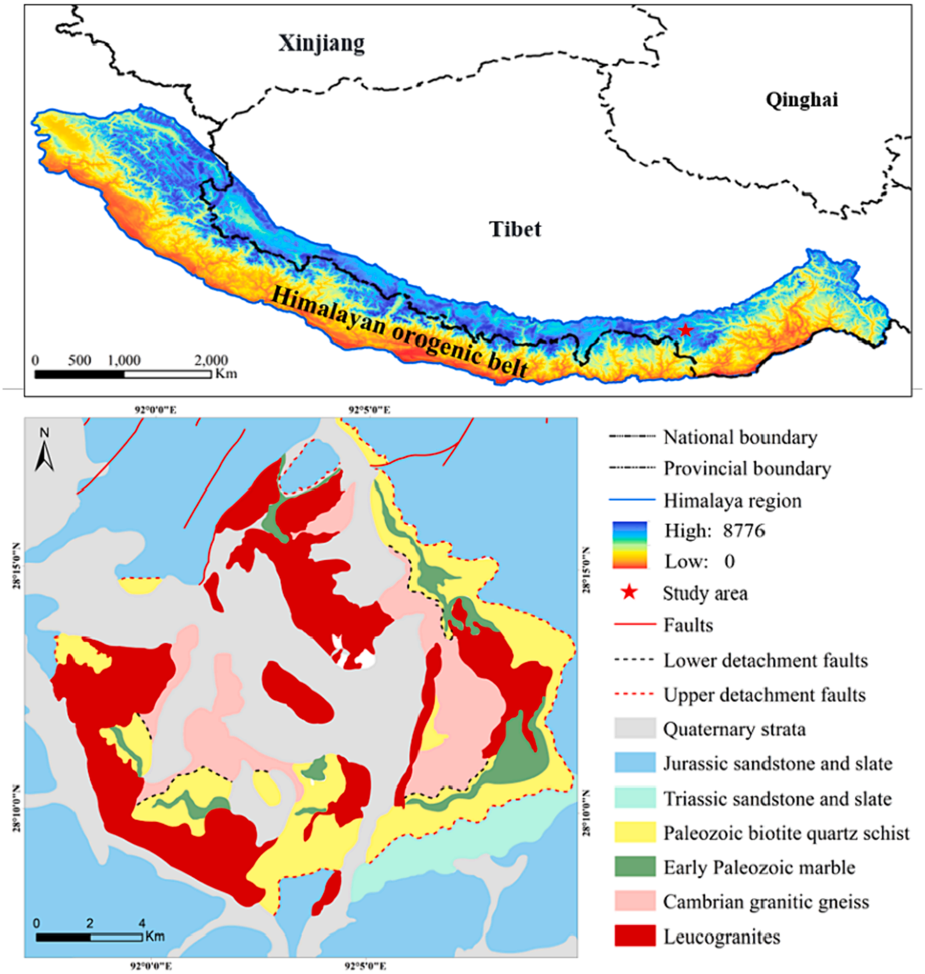


Fig. 1. Simplified geological map of Cuonadong dome, southern Tibet, China (after from Cao et al., 2021).

Table 1
Typical lithological features and major ore minerals of Cuonadong dome.

Class name	Exposure area	Major ore minerals
Jurassic sandstone and slate	Situated around the dome	phyllite, sandstone, calcareous slate
Early Paleozoic marble	Located in the dome mantle part, banded and undergone intense silicification	calcite, dolomite
Triassic sandstone and slate	Located at the bottom of the dome, smaller area	feldspar, shale, slate, fine sandstone, quartz sandstone, sandy slate
Cambrian granitic gneiss	Located in the core of the dome and associated with leucogranite, Gneiss tectonics	potassium feldspar, plagioclase, quartz, biotite, muscovite, et al.
Paleozoic biotite quartz schist	Located in the mantle part of the dome, faceted distribution characteristics	strongly metamorphosed and deformed quartz diamictite schist interbedded with mylonitized carbonate rocks
Quaternary Strata	Covered the entire dome and weathered highly	gravel, clay
Himalayan leucogranite	Located in the core of the dome, mostly in the form of small rock lines or veins	quartz, plagioclase feldspar, potassium feldspar, muscovite, biotite, et al.

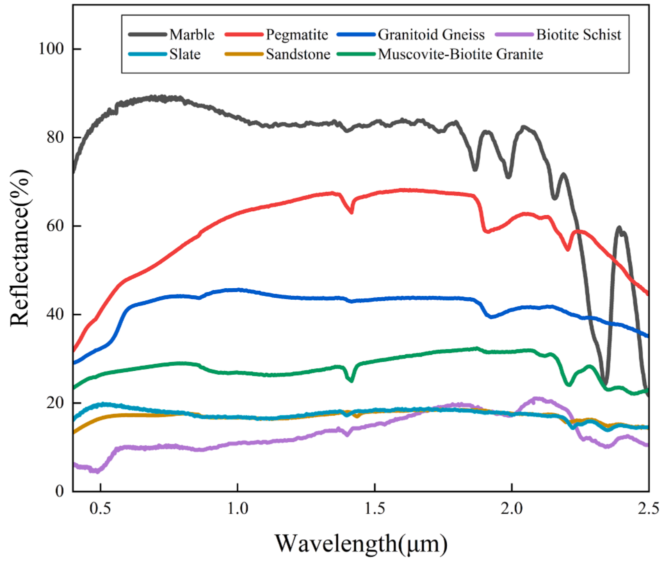


Fig. 2. Spectral curves of several rocks in VNIR.

4.1. Data preprocessing

It is essential for some preprocessing of raw RS images. This includes atmospheric correction, orthorectification, spatial registration, image stitching, and cropping. Atmospheric correction of the S2B image uses the Sentinel Application Platform (SNAP) software and the Sen2cor software package (Main-Knorn et al., 2017). During the correction process, the 10th band, which is the water vapor absorption band, is eliminated. Band 2 with a 10 m spatial resolution is used as the reference

Table 2
Main parameters of the S2B multispectral sensors and GF-5 hyperspectral sensors.

Satellite payloads	Multispectral sensors	Hyperspectral sensors
	S2B	GF-5
Nations	Europe	China
Launch time	2015.7.23	2018.5.9
Spectral range/μm	0.4–2.4	0.4–2.5
Number of bands	13	330
Spectral resolution/nm	–	5(VNIR)/10(SWIR)
Spatial resolutions of used bands/m	10	30
Swath width/km	290	60

for bilinear interpolation. The S2B data is resampled to 10 m. GF-5 HSI data preprocessing is performed in ENVI software (Cooley et al., 2002), using 30 m DEM data (https://eop-cfi.esa.int/index.php/docs-and-mission-data/dem) for orthorectification of GF-5 data, and the Fast Line-of-Sight Atmosphere (FLAASH) module is used for atmospheric correction. Furthermore, we remove bands (151–154, 192–204, 206, 246–264, 314–316, 326–330) from the GF-5 image because they are affected by the atmosphere and water. The remaining 285 bands are used for the experiments. The GF-5 data are registered using multispectral data (S2B data) as a reference image to ensure a spatial error of less than one pixel. Finally, the data sets obtained in the study are clipped and spliced for fusion. The size of the GF-5 dataset in the study area is 805 × 825, and the size of the S2B dataset is 2415 × 2475. Subsequently, we fuse these two RS images with different resolutions obtained by preprocessing as basic data and acquire the fused data for training the network.

4.2. Data fusion

SFIM (Liu, 2000), a classical pixel-wise fusion method, improves the distortion of image spectral properties. This is a straightforward spectral preservation fusion technique with reduced solar radiation and a land surface reflection model. Using the ratio between the high-resolution (HR) image and its low-pass filtered version, the spatial features of the co-registered low-resolution (LR) image can be modified without changing its spectral characteristics, as shown in Fig. 4. More specifically, by multiplying the texture of the HR image and the spectral reflectance of the original LR image, fused high spatial and spectral resolution image can be obtained, which retain most of the spectral information of the LR image. Thus, SFIM enhances the spatial clarity of the original LR image while reducing spectral aberrations.

The steps of the SFIM fusion algorithm are as follows (An and Shi, 2014).

1) Find the degraded version of HR image by using an averaging filter;
2) Compute the ratio between the HR image and the degraded version;
3) Obtain the fused image by multiplying the LR image with the ratio.

Here, the HR image represents the multispectral data (i.e., S2B image), and the LR image represents the hyperspectral data (i.e., GF-5 image). Generally, the SFIM algorithm is a ratio method that introduces spatial details without changing the spectral information and is defined as

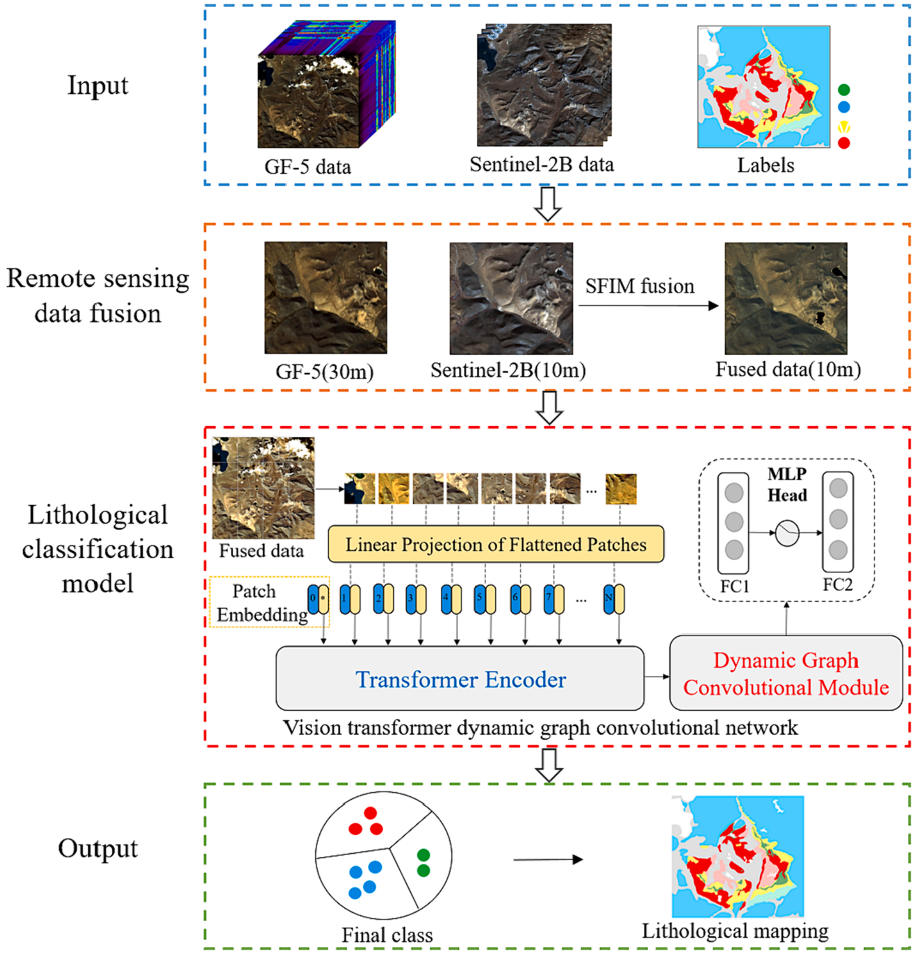


Fig.3. Basic workflow of this paper.

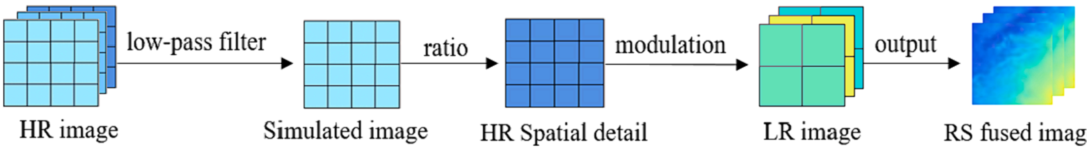


Fig. 4. The process schematic of the SFIM fusion algorithm.

$$\text{IMAGE}_{SFIM} = \frac{\text{IMAGE}'_{low} \times \text{IMAGE}_{high}}{\text{IMAGE}_{mean}} \tag{1}$$

$$\text{IMAGE}'_{low} = \text{Interpolation}\ (\text{IMAGE}_{low}) \tag{2}$$

where IMAGE_{high} is an HR image, and IMAGE_{mean} is a simulated LR image derived from IMAGE_{high} using a low-pass averaging filter. IMAGE'_{low} is the resampled LR image IMAGE_{low}, which has the same pixel size as the HR image. Otherwise, the spatial details of the HR image cannot be completely incorporated into the LR image (Liu, 2000). In this study, we use bilinear interpolation method for the GF-5 data to obtain the same spatial resolution as the S2B data, which is specified in equation (2).

The filter kernel size is based on the spatial ratio between the HR and LR images. For instance, to fuse a 30 m resolution GF-5 image with a 10 m resolution S2B image, the minimum smoothing filter kernel size for calculating the local mean of the S2B image pixels is 3 × 3 defined as

$$\frac{1}{9}\begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$$

Moreover, the ratio in equation (1) cancels the spectral and topographical contrast of the HR image but retains its HR texture information (Li et al., 2004). The results were unrelated to the spectral characteristics of the HR image used for modulation fusion. Therefore, SFIM is dependent on the spectral information and context of the original LR image (Behnia, 2005).

4.3. Vision transformer dynamic graph convolutional network

The framework of the proposed algorithm is illustrated in Fig. 5. First, the ViT-DGCN uses principal component analysis (PCA) for dimensionality reduction. Second, the fused image is divided into patches. The patch size is set to 4 × 4, and the total number of patches is 373, 272. The patch embedding is then sent to the transformer to capture spatial sequence relationships. Next, the sequential features are input into another dynamic GCN module. Finally, the obtained structural representations are used to output the classification results through a fully connected layer. In this work, the proposed ViT-DGCN model has two key components, a transformer and a dynamic GCN module, which are elaborated as follows.

4.3.1. Transformer encoder

Fig. 6 shows the detailed layers of the transformer encoder.

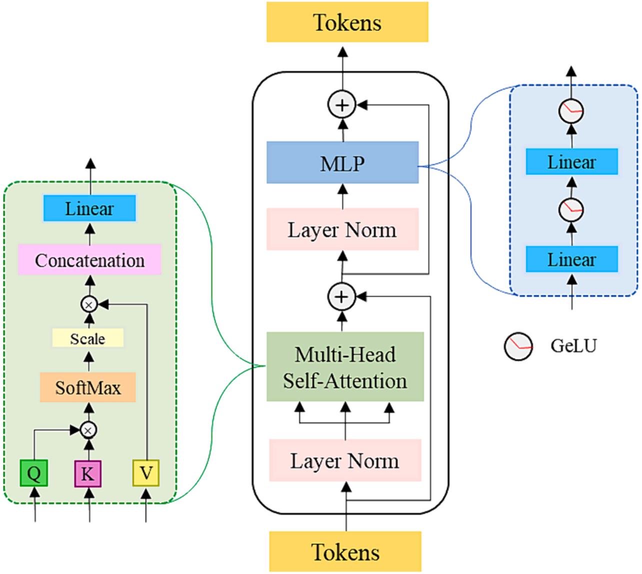


Fig. 6. Network structure of transformer encoder module.

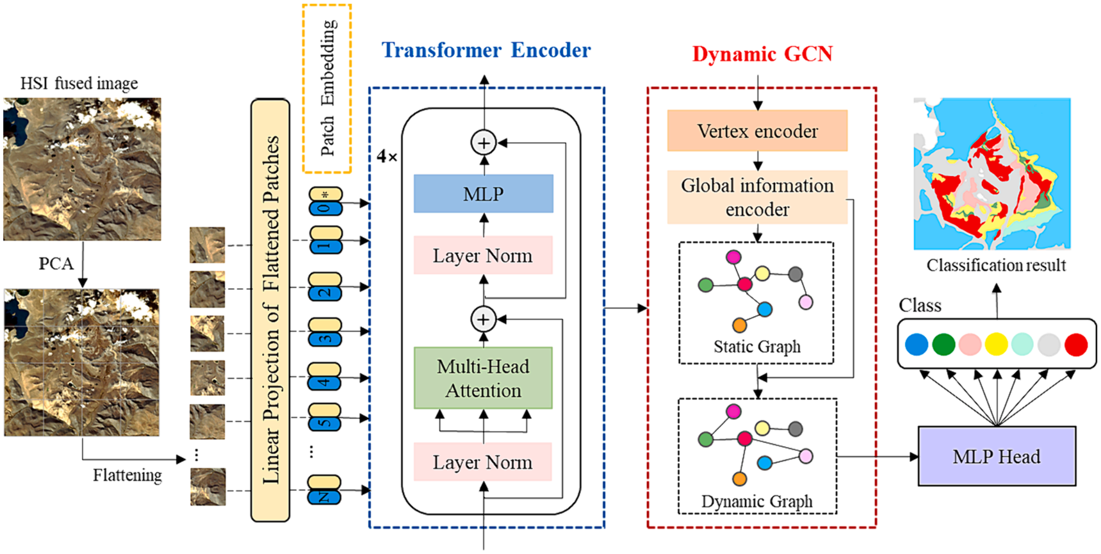


Fig. 5. A brief framework of the proposed ViT-DGCN.

Considering the pixel-level and small sample classification tasks, we use only four blocks of the encoder. Owing to the fixed kernel size and limited number of layers in the CNN, it is challenging to capture long-range dependencies from HSI data, which ignores the sequence information of the inputs (Wang and Tax, 2016). By utilizing the self-attention mechanism, it is possible to examine the correlation between every pair of patches and obtain a better representation of internal features with less reliance on external information. The transformer encoder block mainly includes the layer normalization (LN), MHSA mechanism, and multilayer perceptron (MLP) layer.

Unlike the classic transformer structure, we use the encoder module of a transformer without using the decoder module. In addition, we remove the position encoding to reduce the computational effort and speed up the computation. For the flattened patches, a linear projection is performed to obtain tokens, and then all tokens are fed to the MHSA mechanism. When extracting vector sequence features, the MHSA splits a self-attentive module into multiple heads to focus on different patches of the sequence signal in parallel and projects their joint output, which can capture richer feature information. The attention operation of each head is defined as follows

$$\text{Attention }(Q, K, V) = \text{Softmax }(\frac{QK^T}{\sqrt{d_k}})V \tag{3}$$

where queries $Q = XW^q$, keys $K = XW^k$, values $V = XW^v$ are the linear transformation to input X, $W \in \mathbb{R}^{d_k \times 1}$ denotes the linear projection matrix, and d_k represents the dimension of the vector in K.

This is repeated several times to generate Q, K, and V, and then these results are concatenated. This process is referred to as the MHSA.

$$\text{Multi - head }(Q, K, V) = \text{Concat }(\text{head}_1, \text{ head}_2, \ldots, \text{ head}_\text{h})W^\text{o} \tag{4}$$

$$\text{head}_i = \text{Attention }(Q, K, V) \tag{5}$$

where W^o is a parameter matrix and h denotes the number of heads. The MHSA mechanism is beneficial for increasing the regularization and robustness of the model.

Subsequently, it is delivered to the MLP layer to further transform features learned in the MHSA mechanism. Here, the MLP consists of two fully connected layers separated by the Gaussian error linear unit (GELU) nonlinearity. The GELU activation function is calculated as follows

$$\text{GELU} = x\Phi(x) = x \cdot \frac{1}{2}\left[1 + erf(\frac{x}{\sqrt{2}})\right] \tag{6}$$

$$erf(x) = \int_0^x e^{-t^2} dt \tag{7}$$

where $\Phi(x)$ indicates the standard Gaussian cumulative distribution function.

The LN always precedes the MLP, both of which decrease the training time by normalizing the neurons and alleviating the problem of gradient disappearance or explosion. The GELU activation function introduces stochastic regularization, which enables the network to converge faster and improves its generalization ability. Furthermore, an identity MLP layer is added to increase the nonlinear mapping capability of the model. In conclusion, the transformer has the advantage of obtaining sequence relationships over patch embedding, which aims to capture the interactions between all patches by encoding global contextual information.

4.3.2. Dynamic graph convolutional module

To further mine the dynamic graph features among the data and compensate for the fact that the transformer lacks some of the inductive biases inherent to the CNN, we introduce dynamic GCN, which can efficiently process non-Euclidean structure data and have the ability to learn topological information (Liu et al., 2022). The designed dynamic graph convolutional network consists of two main modules, the vertex encode module (VEM) for obtaining vertex features and the dynamic graph convolutional layer for adaptively acquiring graph structure features. As shown in Fig. 7, the VEM first calculates the category-specific activation maps and passes them into the dynamic graph convolution layer to acquire the global topology information.

VEM refers to vertex-wise encoding and pooling for graph classification. Its objective is to generate a set of vertex feature that represent the content related to different labels from an input feature map $\boldsymbol{p}$. The first step of VEM is to compute activation maps $\boldsymbol{M} = \{\text{m}_1, \text{m}_2, \ldots, \text{m}_c\}_{c=1}^{C}$ specific to each category, and C is the number of categories. These maps are then utilized to transform the input feature map into sequence representations $\boldsymbol{N} = \{n_1, n_2, \ldots, n_c\}_{c=1}^{C}$. Formula (8) gives its calculation process.

$$n_c = \text{m}_c^T \boldsymbol{p}' \tag{8}$$

where m_c^T denotes the transposed weight of the cth activation map. Vertex information can then be used to selectively aggregate relevant category-specific features.

We develop a novel dynamic GCN that adaptively extracts the structural features of aggregated neighbor nodes. The objective of the dynamic GCN layer is to update the value of $\boldsymbol{N}$ using a learnable weight matrix $\boldsymbol{W}$ and adjacency matrix $\boldsymbol{A}$ for state updates.

In particular, the first GCN layer is a regular graph convolution operation. The process can be formulated as

$$\boldsymbol{N}' = \delta(\boldsymbol{ANW}) \tag{9}$$

where the adjacency matrix $\boldsymbol{A}$ records the relations between the features of each node.

$\delta(\cdot)$ denotes an activation function LeakyReLU. It can be defined as

$$\text{LeakyReLU}(x) = \begin{cases} x, & x > 0 \\ \eta x & x < 0 \end{cases} \tag{10}$$

where the value of η is 0.01 and it represents a negative slope coefficient.

We then introduce the adjacency matrix $\boldsymbol{A}'$ to update the node $\boldsymbol{N}'$ in the next layer. The $\boldsymbol{A}$ in the previous layer is fixed, while $\boldsymbol{A}'$ can be dynamically updated with change in the input features. We apply the fully connected network to transform the feature vectors $\boldsymbol{F}$, $\boldsymbol{G}$, and $\boldsymbol{H}$ to obtain new representations.

$$\alpha_{ji} = \frac{\exp(\boldsymbol{F}_i \cdot \boldsymbol{G}_j)}{\sum_i^S \exp(\boldsymbol{F}_i \cdot \boldsymbol{G}_j)} \tag{11}$$

$$\boldsymbol{A}' = \beta \sum_{i=1}^{S} (\alpha_{ji} \boldsymbol{H}_i) + \boldsymbol{N}_j \tag{12}$$

where β is the learnable parameter. Because every patch has a different $\boldsymbol{A}$, it enhances its representative ability and decreases the risk of overfitting caused by a static graph. Formally, the output $\boldsymbol{N}''$ of the dynamic graph convolution layer is expressed as

$$\boldsymbol{N}'' = \delta(\boldsymbol{A}'\boldsymbol{N}'\boldsymbol{W}') \tag{13}$$

where $\boldsymbol{W}'$ represents learnable parameters. Our dynamic GCN can improve content-aware category representations thanks to its dynamic graph convolutional layer.

To overcome the limitations of the local context in deep neural networks, we utilize a transformer-based architecture to extract image features. The proposed ViT-DGCN can fully learn global sequence features and topological graph structure information, demonstrating its powerful feature extraction ability.

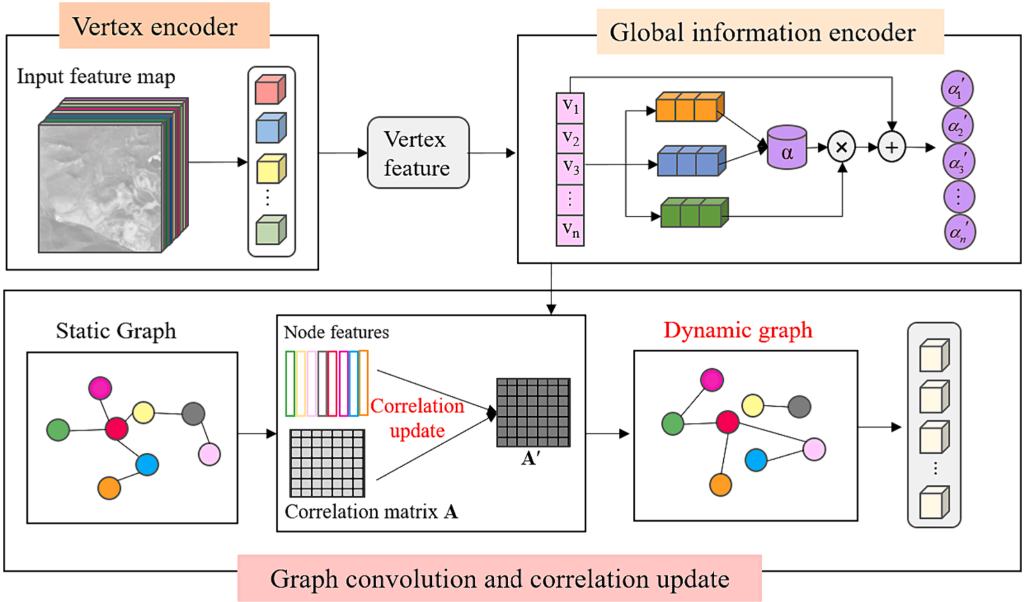


Fig. 7. Network structure of dynamic GCN module.

5. Experiments

In this section, we describe the experimental setup, including the comparison algorithms, parameter settings, and analysis of ViT-DGCN. 1 % of samples are randomly selected as training set and the remaining are regarded as testing set in our classification experiments. Table 3 shows the training samples selected for each category. of Cuonadong dataset. There are 7 lithological classes in the Cuonadong dataset. The water and clouds in the images are masked and regarded as class 8.

5.1. *Comparative methods*

Seven classic or recently proposed supervised classification methods are used comparison methods, including a traditional and typical ML algorithm RF (Breiman, 2001) and six DL-based methods, namely spectral–spatial residual network (SSRN) (Zhong et al., 2018), deep feature fusion network (DFFN) (Song et al., 2018), deformable HSI classification networks (DHCNet) (Zhu et al., 2018), fast dense spectral–spatial convolution network (FDSSC) (Wang et al., 2018), HSI classification with transformers SpectralFormer (Hong et al., 2021) and fast dynamic GCN and CNN (FDGC) (Liu et al., 2022). The DL methods code can be found at (https://github.com/Candy-CY/Hyperspectral-Image-Classification-Models). Note that RF uses pixel-based samples, DFFN, DHCNet, SSRN, FDSSC, SpectralFormer, FDGC, and the proposed ViT-DGCN use patch-based samples. The overall accuracy (OA), kappa coefficient, and accuracy of each class are calculated to quantitatively measure classification effectiveness. For a fair comparison, the experimental results are obtained five times and averaged.

5.2. *Experimental settings*

The setup of network parameters has a considerable impact on the performance of deep networks. For the parameters inside the proposed model, the token embedding dimension t is 64, the self-attention in the MHSA has 4 heads, and the number of neurons for the hidden layer of the model is 64, with an increased ratio of 4 for the MLP hidden layer. The number of training iterations is 100. The activation function is LeakyReLU and the optimizer is Adam with label smoothing. The detailed parameters of the ViT-DGCN model are listed in Table 4. Our experiments were performed using Python-3.8.13 and PyTorch-1.10.0.

5.3. *Parameter analysis*

The accuracy and efficiency of the ViT-DGCN model are significantly influenced by certain hyperparameters, including the model depth, batch size, input patch size, and learning rate. The optimal values of these hyperparameters are determined using a controlled variable method. Thus, we examine the parameters affecting the training performance of the model.

1) Dimensions of HSI: Hyperspectral data have hundreds of bands, and processing high-dimensional hyperspectral images without data dimensionality reduction has the potential to increase the computation. Therefore, the dimensional setting value of PCA dimensionality reduction is a key parameter. Fig. 8 shows the performance of the ViT-DGCN for different dimensions. The results show that as the number of dimensions increases from 5 to 50, the OA reaches nearly 97 % at 20 dimensions achieving a relatively good result, and reaches

Table 3
Randomly select the sample size for each category.

Class label	Class name	Area (km^2)	Training samples: 1 %
1	Jurassic sandstone and slate	286.6	28,558
2	Early Paleozoic marble	11.5	1146
3	Triassic sandstone and slate	13.0	1297
4	Cambrian granitic gneiss	29.5	2955
5	Paleozoic biotite quartz schist	47.8	4149
6	Quaternary Strata	125.9	12,557
7	Himalayan leucogranite	62.8	6201
8	Water and clouds	26.2	2617

Table 4
The detailed parameters of the ViT-DGCN model. Note: S, T and B denote the length, width and number of bands of the input dataset, respectively.

Layers	ViT-DGCN	
Input	S × T × B	35 × 27 × 27
Tokenization	Patch embedding	64
Transformer block × 4	Attention head	4
	MLP ratio	4
Dynamic GCN	Graph convolution layer	64
	Fully connection layer	64
Output	Softmax	1 × C

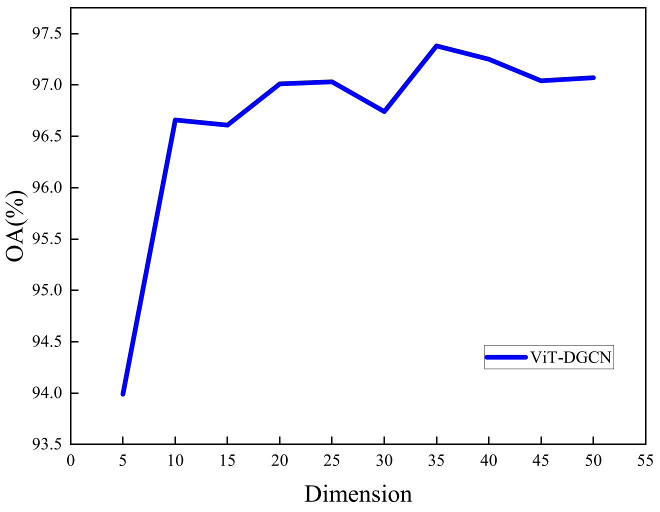


Fig. 8. Performance comparison of the ViT-DGCN model with different dimensions.

a maximum of 97.38 % at 35 dimensions. Therefore, the ViT-DGCN model with 35 dimensions is set owing to its excellent classification accuracy.

2) Depth of the ViT-DGCN: Network depth plays a crucial role in determining the complexity of the transformer network. A higher depth can result in a larger number of parameters and increased complexity, leading to potential challenges such as computational burden and overfitting. On the other hand, a lower depth reduces complexity and offers advantages in terms of computational resources, but may result in underfitting issues. Therefore, determining an appropriate depth configuration can effectively optimize the trade-off between classification accuracy and computational efficiency for lithological mapping tasks. As shown in Fig. 9, as the depth of the network increases, the training time increases proportionally. The OA first increases and then decreases, indicating that as the complexity of the model increases, the model requires more data and time to adjust the weights. However, HSI data are limited, hindering the performance of the algorithm. Therefore, considering the time efficiency and classification effect, the network depth is set to 4.
3) Learning rate: As the crucial hyperparameter, the learning rate controls the convergence speed of the DL model. A suitable learning rate can contribute to stable and rapid convergence of the model.

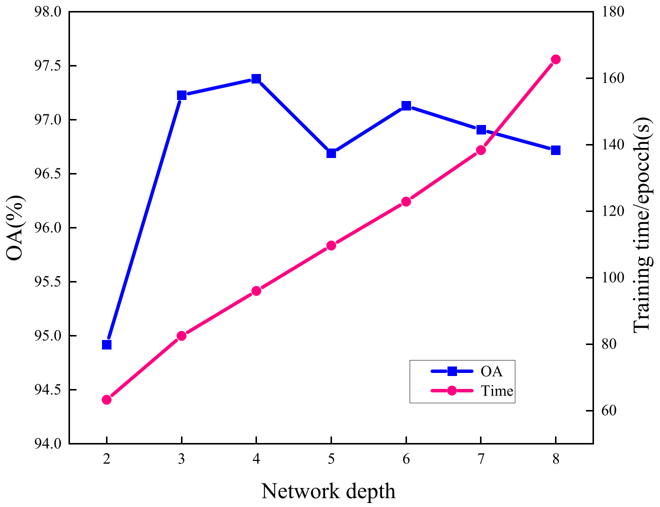


Fig. 9. Performance comparison of the ViT-DGCN model with different network depth.

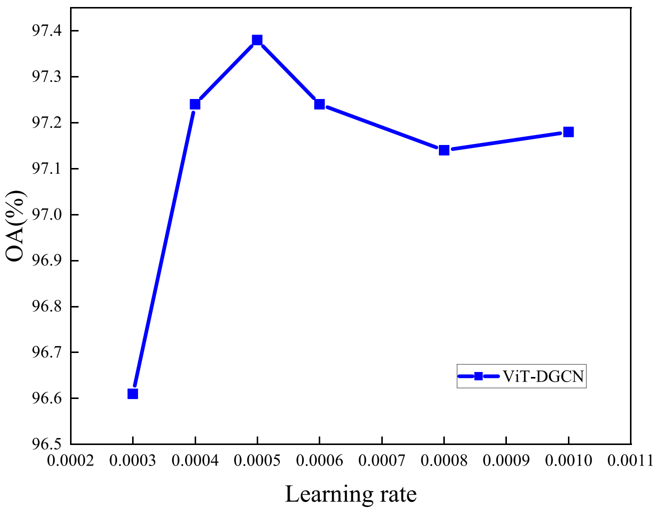


Fig. 10. Performance comparison of the ViT-DGCN model under different learning rates.

Fig. 10 shows the classification accuracy of ViT-DGCN for different learning rates. By comparison, 0.0005, with the highest accuracy, is selected as the learning rate value for ViT-DGCN.

4) Training Batch size: Batch training is a crucial operation in DL that allows for the parallel processing of multiple input data, improving memory utilization and network optimization capability. To determine the appropriate batch size, we conduct experiments to determine and compare the accuracy and efficiency of the classification. Fig. 11 presents the classification results of the ViT-DGCN model for batch sizes ranging from 64 to 512. From the perspective of the classification accuracy, we can observe that the highest OA and reasonable training time is achieved when a batch size of 128. Therefore, 128 is selected as the discounted batch size.
5) Input Patch size: The input size of the HSI patches plays a vital role in determining the amount of information that can be effectively utilized by the DL model. If the patch size is excessively small, it may lead to insufficient valid information. Conversely, using a large patch size would increase computational costs and introduce potential distractions. To find the appropriate patch size, Fig. 12 shows the performance of the ViT-DGCN under different patch size ranging from 15 × 15 to 35 × 35. The results indicate that a patch size of 27 × 27 achieve the best accuracy, while maintaining an acceptable

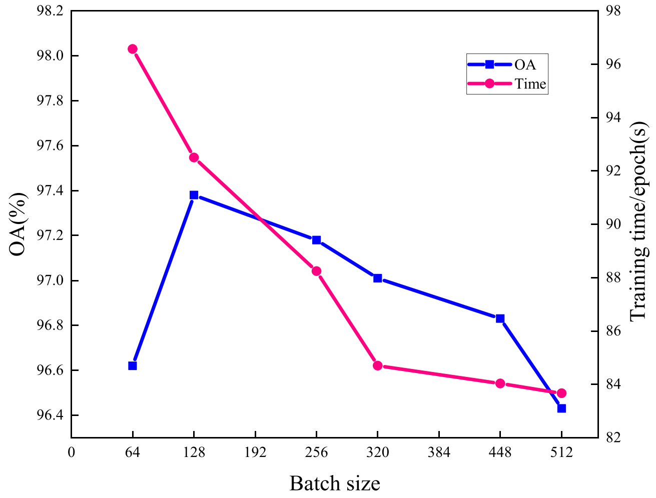


Fig. 11. Performance comparison of the ViT-DGCN model with different batch size.

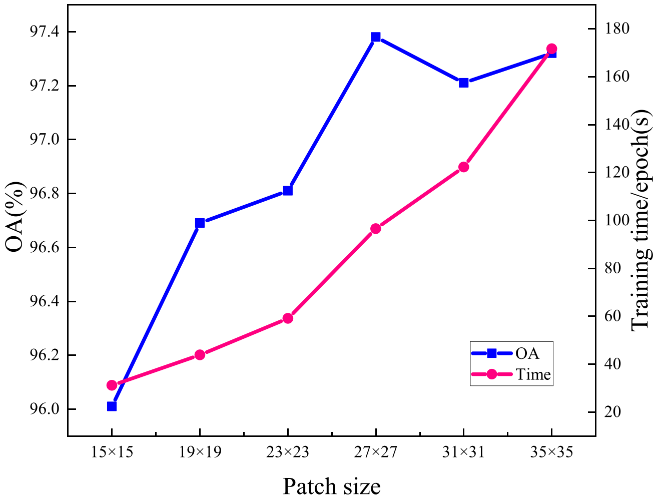


Fig. 12. Performance comparison of the ViT-DGCN model with different patch size.

time cost. Consequently, a 27 × 27 patch size is selected as the final input.

6. Results

In this section, we first evaluate the impact of integrating GF-5 and S2B data. Next, we quantitatively assess and interpret the practical classification results of the algorithms using the Cuonadong dataset. Finally, we demonstrate the visualized lithological classification maps and analyzed the distribution of the rock units.

6.1. Fusion of GF-5 and S2B data

Fig. 13 presents the visualized RGB false-color maps and magnified part of GF-5 data, S2B data, and the fused data, which demonstrate the spatial enhancement achieved by the fused data compared to the original datasets. The performance of the downstream classification application is selected as the evaluation standard to verify the fusion effect. It has usually been found that the RF method has better classification results than other widely used ML algorithms in lithological mapping (Shayeganpour et al., 2021; Xi et al., 2022). The RF and ViT-DGCN methods are chosen to compare the effects of fusion before and after fusion.

Table 5 and Table 6 present the quantitative assessment results of the GF-5 and S2B image and fused image. Bold type indicates the best result. It is clear that the classification accuracy after fusion increased by about 10 % compared with the original data. There is a significant improvement in the classification accuracy after the fusion of Class 2 Early Paleozoic marble and Class 3 Triassic sandstone and slate, which demonstrates the effectiveness of SFIM fusion.

Table 5
Fusion results comparison of RF classification.

Class label	S2B	GF-5	SFIM_GF-5 + S2B
1	94.12 ± 0.07	93.57 ± 0.32	**96.18 ± 0.06**
2	9.72 ± 0.51	7.45 ± 1.04	**30.56 ± 0.72**
3	12.84 ± 0.35	15.10 ± 0.23	**35.38 ± 0.77**
4	37.13 ± 0.57	37.78 ± 1.91	**59.99 ± 0.89**
5	16.35 ± 0.44	13.82 ± 0.95	**40.93 ± 0.49**
6	66.86 ± 0.16	71.00 ± 0.61	**80.75 ± 0.08**
7	46.46 ± 0.32	47.20 ± 0.35	**66.67 ± 0.48**
OA (%)	70.71 ± 0.12	71.85 ± 0.66	**80.95 ± 0.07**
kappa × 100	53.93 ± 0.13	57.09 ± 0.39	**70.76 ± 0.12**

Table 6
Fusion results comparison of ViT-DGCN classification.

Class label	S2B	GF-5	SFIM_GF-5 + S2B
1	93.46 ± 0.11	96.45 ± 0.04	**98.83 ± 0.23**
2	66.24 ± 1.45	71.27 ± 1.85	**88.48 ± 0.11**
3	82.38 ± 0.15	89.30 ± 0.35	**95.41 ± 0.79**
4	77.38 ± 0.05	87.45 ± 0.58	**95.19 ± 0.23**
5	76.39 ± 1.26	78.58 ± 0.24	**92.89 ± 0.33**
6	88.00 ± 0.18	90.20 ± 0.30	**96.37 ± 0.35**
7	81.64 ± 0.55	84.47 ± 1.42	**95.52 ± 0.30**
OA (%)	88.33 ± 0.18	91.49 ± 0.13	**97.15 ± 0.20**
kappa × 100	82.61 ± 0.28	87.39 ± 0.55	**95.81 ± 0.28**

The RF and ViT-DGCN classification results and magnified boundaries before and after fusion are shown in Fig. 14 and Fig. 15. It is obvious that the misclassification is greatly reduced, and the geological boundaries are clearer. The visual analysis also indicates that fusion process is effective and beneficial for lithological discrimination and classification.

6.2. Classification results

Table 7 lists the evaluation metrics comparison of seven methods for the Cuonadong dataset. The optimal value for each indicator is in bold. RF yields lower accuracy compared to other DL-based methods, which may result from ignoring the spatial characteristics of the neighboring data. Also, due to the intra-class variability and spectral variability phenomena of HSI, RF cannot effectively distinguish these lithological categories.

The distribution of rocks and minerals in the Cuonadong dome is highly unbalanced. For class 1 Jurassic sandstone and slate with wide coverage, the recognition accuracy based on both RF and DL-based models reach more than 93 % with no noticeable improvements

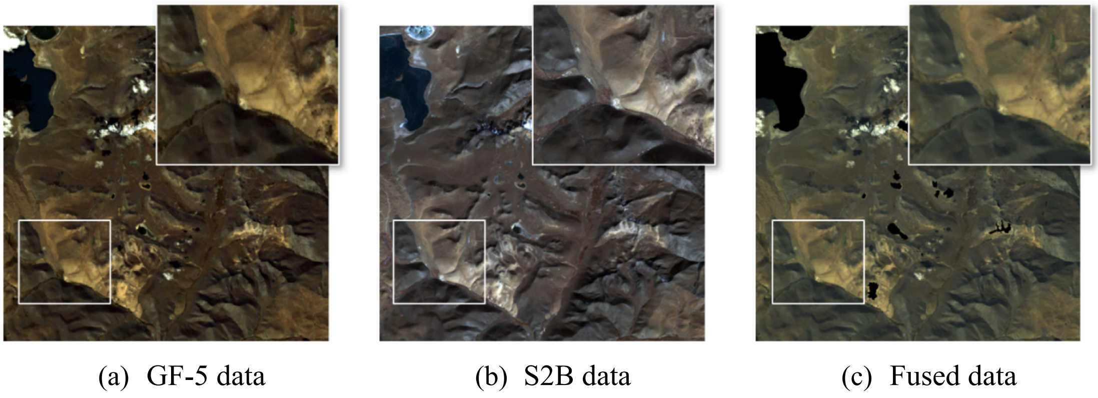


Fig. 13. The visualized RGB false-color map and magnified part of different data.

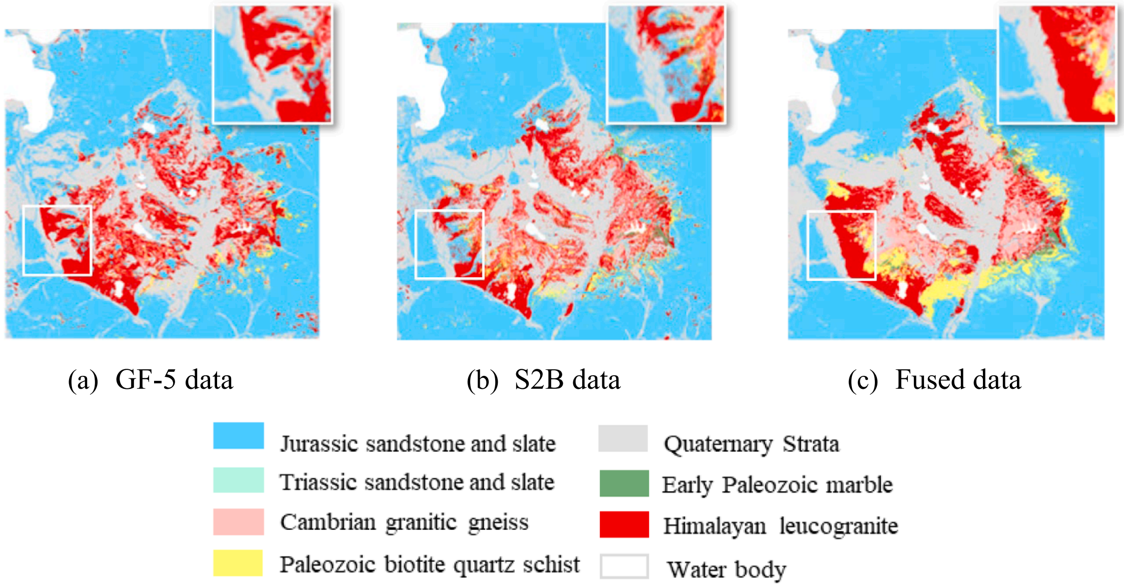

(a) GF-5 data (b) S2B data (c) Fused data

Fig.14. Classification results and magnified boundaries of lithological units of different data and RF classifier.

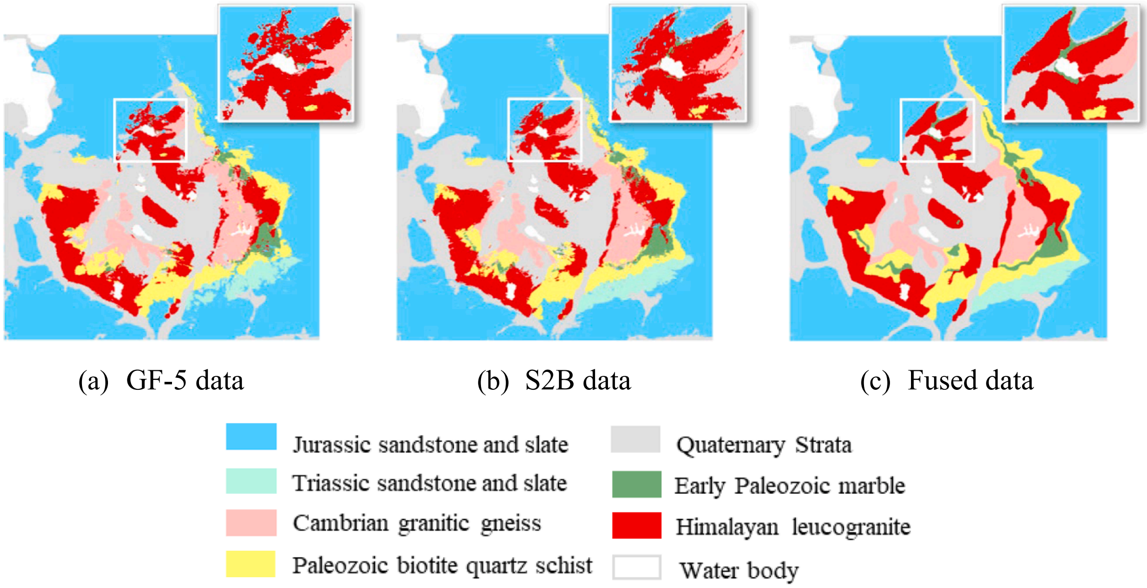

(a) GF-5 data (b) S2B data (c) Fused data

Fig.15. Classification results and magnified boundaries of lithological units of different data and ViT-DGCN model.

because the apparent differences in mineral content are easily identified. However, the proposed method, with excellent feature extraction capability, shows the best accuracy for class 1 Jurassic sandstone and slate.

Nevertheless, as a lithological category with little distribution of lithologic units, such as class 2 Early Paleozoic marble, and class 3 Triassic sandstone and slate, some methods including DFFN, DHCNet, and SSRN, have obtained relatively low-class accuracies. First, the spatial coverage of the two lithological units is extremely small with a banded and linear distribution leading to poor classification. Furthermore, these lithological classes are affected by severe weathering, which results in a considerable loss of spectral feature information.

SpectralFormer is an HSI classification method with transformers that outperforms than classic backbone CNN networks, such as DFFN, DHCNet, and SSRN. The FDSSC and FDGC achieved good accuracies in class 2 and 3, and produced encouraging results, with an OA of approximately 93 %. The FDSSC framework uses different convolutional kernel sizes to mine high-level spectral–spatial information, which explains its relatively good accuracy. The good results of the FDGC can be explained by the positive influence of the specific network design integrating the dynamic GCN and CNN. Despite this, our method exhibited the best performance for the recognition of small rock masses. It can be seen that our model outperforms other DL methods except FDGC in terms of training time, but our model has the highest accuracy. We believe it is worth spending extra time to get better performance.

Compared with all models, the proposed ViT-DGCN achieves the highest classification accuracy, which indicates the usefulness of the incorporation of transformer and dynamic GCN. In particular, our ViT-DGCN method achieves a 96 % recognition rate for class 7 Himalayan leucogranite with a noticeable improvement over the other methods, providing technical ideas for further targeting of objective lithology for rare metal exploration. Our proposed model organically combines transformer with dynamic GCN to utilize the topology and sequence features. By extracting more discriminative and representative features,

Table 7
Comparison of classification results among different models.

Class label	RF	DFFN	DHCNet	SSRN	FDSSC	Spectral Former	FDGC	ViT-DGCN
1	96.18 ±0.06	93.44 ±1.37	94.05 ±1.35	96.8 ±0.67	96.22 ±0.42	96.33 ±0.78	97.5 ±0.17	**98.83 ±0.23**
2	30.56 ±0.72	43.63 ±4.50	49.69 ±6.35	58.6 ±9.71	74.47 ±8.25	63.62 ±5.38	74.08 ±5.45	**88.48 ±0.11**
3	35.38 ±0.77	60.35 ±5.81	68.62 ±3.61	80.56 ±1.95	87.25 ±4.17	83.95 ±5.00	89.17 ±1.35	**95.41 ±0.79**
4	59.99 ±0.89	67.28 ±6.53	65.31 ±11.0	75.62 ±4.64	85.6 ±5.07	84.01 ±0.52	86.86 ±0.58	**95.19 ±0.23**
5	40.93 ±0.48	56.5 ±1.72	63.26 ±5.79	72.19 ±3.62	82.97 ±4.45	76.55 ±0.83	82.52 ±1.08	**92.89 ±0.33**
6	80.75 ±0.08	85.21 ±3.65	83.5 ±2.35	87.43 ±4.80	93.56 ±0.89	90.28 ±0.30	91.68 ±0.83	**96.37 ±0.35**
7	66.6 ±0.06	72.87 ±3.14	75.33 ±1.21	84.23 ±2.05	88.28 ±1.17	84.89 ±0.99	88.11 ±0.32	**95.52 ±0.30**
OA (%)	80.95 ±0.48	83.72 ±0.55	84.87 ±1.12	87.61 ±3.45	92.55 ±0.31	90.88 ±0.04	93.39 ±1.13	**97.15 ±0.20**
kappa × 100	70.76 ±0.12	75.82 ±0.86	77.62 ±1.73	81.57 ±5.16	89.59 ±0.46	86.54 ±0.12	90.22 ±1.67	**95.81 ±0.28**
Training time(h)	0.504	10.322	8.059	5.452	5.883	6.478	2.802	3.229

the ViT-DGCN framework demonstrates superior classification accuracy compared with all other models.

6.3. *Lithological mapping*

Fig. 16 shows the lithological mapping results obtained using different methods. Clearly, the proposed ViT-DGCN model provides an excellent visual effect. Visually, the ML-based method RF(a) produces a classification map that shows a lot of noise and misclassification, while the DL-based methods generate classification maps that are smoother and clearer than RF, reflecting the effectiveness of DL-based methods in the application of lithological mapping.

Specifically, to reduce the noise and misclassification phenomenon, the DL-based methods DFFN(b) and SSRN(d) introduce residual learning to extract the spatial-spectral features. DHCNet(c) uses deformable convolution to flexibly extract high-level effective spatial features. Their maps have fewer noise points than RF, and the misclassification phenomenon is somewhat improved. The FDSSC(e) network utilizes an efficient convolution method to reduce high dimensionality. SpectralFormer(f) spectrally learns global sequence information using the transformer. They demonstrate superior performance better than DFFN (b), DHCNet(c), and SSRN(d). FDGC(g) performs relatively well in delineating lithological units and defining their boundaries clearly.

However, based on the lithological classification maps, Himalayan leucogranite is easily wrongly identified as Cambrian granitic gneisses due to their very similar material compositions and spectral characteristics. This can lead to misclassification. Similarly, some of the Early Paleozoic marble units are mistakenly identified as Paleozoic biotite quartz schist, with low recognition of both lithological classes. The irregular ribbon-like and linear distribution of these rocks in the dome mantle makes it difficult to distinguish between them precisely. The classification results for Triassic sandstone slate are relatively unstable and low accurate because it belongs to silicate rocks that are more susceptible to weathering and erosion. As a result, some regions within these lithological units may not be correctly classified.

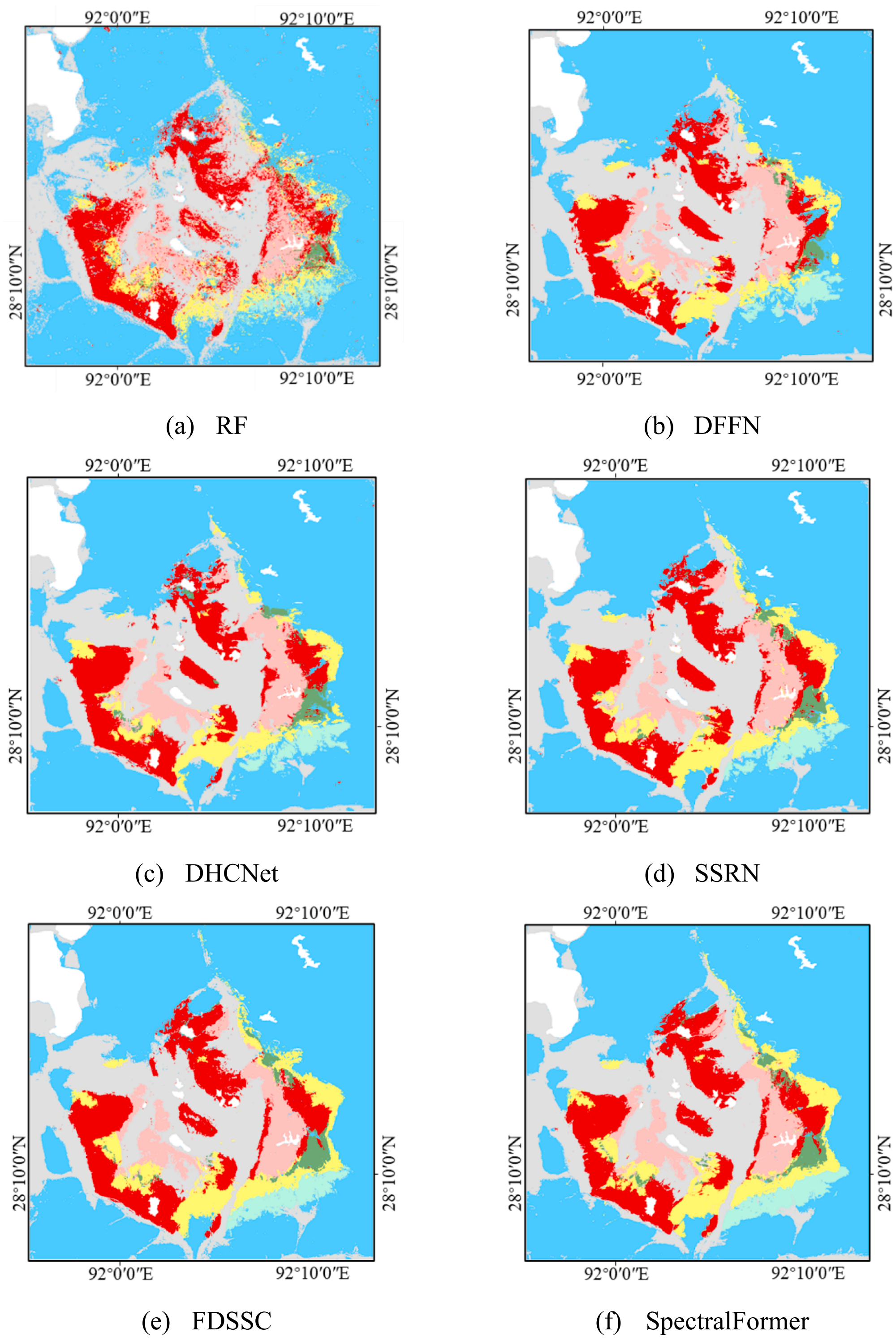
The proposed ViT-DGCN(h) yields an optimal visual result with excellent performance. This map is smooth, and some spatial details are not ignored, such as the Early Paleozoic marble and Triassic sandstone and slate with minor coverage, and the rock boundaries of Himalayan leucogranite are clear and rarely misclassified. To further illustrate the effectiveness of the proposed method for lithological mapping, we have amplified some sections of lithological units in classification maps in Fig. 17. It is evident that ViT-DGCN outperforms other methods with clearer geological boundaries and a minimal number of misclassifications. As expected, the lithological map generated by the ViT-DGCN is more consistent with the ground truth map. Our approach has proven to be superior and rational based on both qualitative and quantitative results.

7. Discussion

Because Himalayan leucogranite has good potential for rare metal mineralization, it can be used as a mineral indicator for rare metal deposits in the Himalayan orogenic belt. Therefore, mapping the spatial distribution of Himalayan leucogranite is considered the primary task in the exploration of rare metal deposits (Cao et al., 2022). Currently, based on the deepening and promotion of RS geological applications, easily accessible RS data are being utilized to better identify geological features, thereby providing accurate directions for mineral exploration. However, it has been challenging to better identify geological features. Rocks are a collection of minerals, and their spectra are essentially a mixture of multiple mineral spectra. Thus, the spectral characteristics of rocks mainly depend on their mineral composition and relative content, as well as being affected by external factors such as surface weathering, rock structure, and surface color (Feng et al., 2018).

Rock bodies exist in certain complex geological environments, and the results of RS lithological identification are affected by environmental factors, especially surface weathering. Weathering not only produces weathering cracks that destroy the structure of the original rock, but also produces alteration minerals that change the mineral composition of the original rock, thereby affecting the physicochemical properties and spectral reflectance characteristics. Therefore, weathering is not conducive to discrimination rock masses. As far as the quantitative results are concerned, it can be seen that the ViT-DGCN model has a slightly poorer classification accuracy for Early Paleozoic marble, owing to the fact that the weathering degree of this class is higher and the spectral features of rocks and minerals are more chaotic.

Limited data sources and complex processing are two major obstacles to the utilization of HSI techniques for lithological mapping and mineral exploration. This study takes advantage of the rich spectral information of HSI data and improves its spatial resolution at the data level, which improves the image environment and image quality, and enhances the reliability of subsequent lithological classification applications. Base on this, the ViT-DGCN model is proposed to facilitate the differentiation of highly similar lithological units. By learning the deep structural features and hidden sequential information of lithologies, the model achieves an impressive classification accuracy of up to 97 %, using only 1 % of the training samples. The above description not only demonstrates that HSI,



(a) RF

(b) DFFN

(c) DHCNet

(d) SSRN

(e) FDSSC

(f) SpectralFormer

Fig. 16. Lithological mapping results of different methods.

with its advantage of extremely high spectral resolution, can sensitively capture such differences in the spectral characteristics of rocks and minerals, but also shows the effectiveness of the proposed model in lithological mapping for mineral exploration. This work will help guide the exploration of similar deposits and provide a technical solution for further exploration of rare metals in the Himalayans.

7.1. Analysis of different training samples

The HSI feature classification within a supervised DL model is a data-driven approach for mining high-level features. Hence, the number of samples utilized during the training stage significantly affects the quality of the DL model. Additionally, proper training samples can conserve computational resources and time without degrading the classification accuracy. As shown in Fig. 18, we perform experiments to

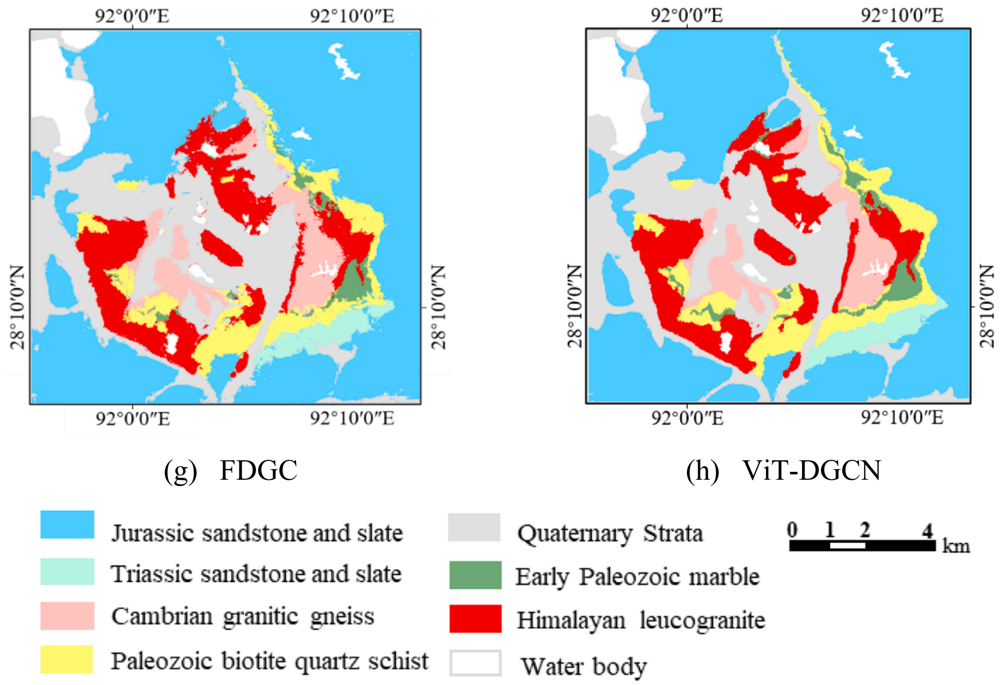


Fig. 16. (*continued*).

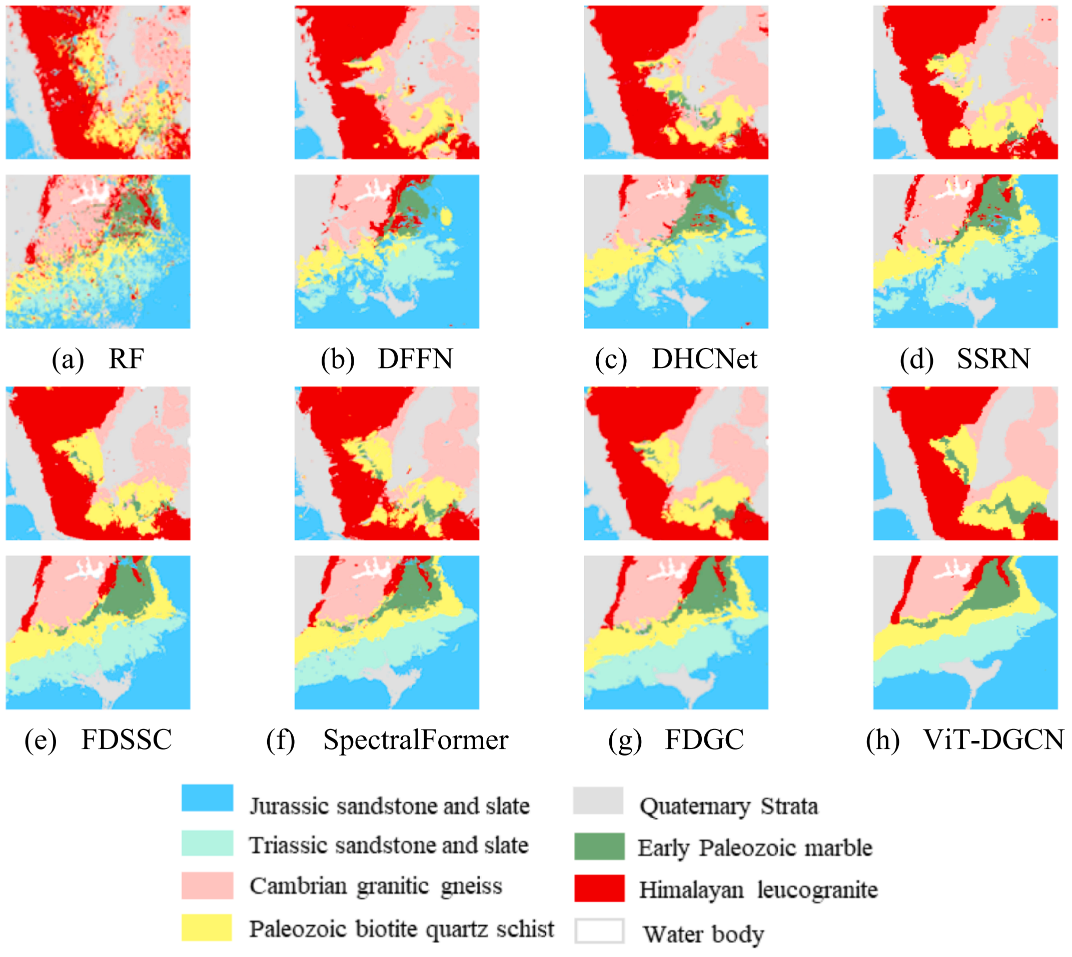


Fig. 17. Magnified parts of lithological Classification results of different models.

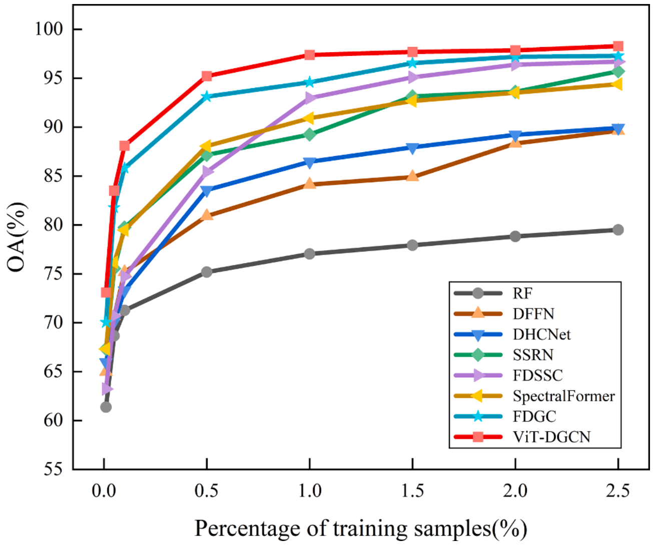


Fig. 18. Performance comparison of the model performance over different percentage of training samples.

evaluate the classification performance of the various models using different training samples. The percentage of training samples is set to 0.01 %, 0.05 %, 0.1 %, 0.5 %, 1 %, 1.5 %, 2 %, and 2.5 %.

It is obvious that the OA of the RF is always lower than that of the DL-based models at different sample proportions because the shallow learning algorithm cannot learn deep abstract knowledge. When the percentage of the training set is large, the ViT-DGCN model displays optimal performance compared to the other comparison models. When the percentage is low, the ViT-DGCN model maintains competitive results. In conclusion, the ViT-DGCN model demonstrates significant improvements when training with limited samples. Owing to its specific and well-designed hybrid structure that combines transformer and dynamic GCN, the proposed ViT-DGCN model exhibits superior performance in terms of OA. These findings indicate that the OA of the proposed method does not show substantial improvement when the training percentage exceeds 1 %. Therefore, using 1 % of the training set during a suitable training period is considered the most appropriate.

7.2. Analysis of model generalization

The generalizability of a model is of great importance when solving practical problems. Model generalization is the ability of a model to suit data from different scenarios. If a model has powerful regularization capability, then it performs excellently on diverse datasets. To assess the generalizability and applicability of the proposed model, experiments are conduct using two public HSI datasets: the Pavia University (PaviaU) satellite dataset and the Salinas airborne dataset (https://www.ehu.eus/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes).

The PaviaU dataset was captured by the Reflective Optics System Imaging Spectrometer (ROSIS) sensor during an aerial survey conducted over Pavia, Northern Italy. This dataset comprised 103 spectral bands with a spatial resolution of 1.3 m and a data size of 610 × 610 pixels. Fig. 19 illustrates the false-color map and ground truth distribution, and the ViT-DGCN classification results of the PaviaU dataset, which consists of 9 classes representing different ground objects.

The Salinas dataset was acquired using the AVIRIS sensor in Salinas Valley, California, which has a spatial resolution of approximately 3.7 m. It is composed of 512 × 217 pixels and 204 spectral bands that cover the range of 400–2500 nm after data preprocessing. Fig. 20 shows the false-color map, ground truth distribution, and the ViT-DGCN classification results of the Salinas dataset, which encompasses 16 classes representing different ground objects.

According to Section 5 of our experimental setup and analysis, 1 % of the training samples are randomly selected from both the PaviaU and Salinas datasets, and the rest are used as testing samples. Table 8 presents the classification results for the two datasets. Evidently, the ViT-DGCN model achieves high classification accuracy, which is comparable to its performance on the Cuonadong datasets. The experimental results obtained from both types of datasets validate the strong generalization ability of the ViT-DGCN model.

8. Conclusion

This study provides an effective and accessible approach for lithological mapping of the Cuonadong dome rare metal deposit using a combination of multi-source RS data fusion and DL method. Hyperspectral and multispectral data fusion technology is first applied in this study area, complementing the advantages of different RS data to provide a high-quality dataset with high spatial and spectral resolution for lithological mapping. A ViT-DGCN model with transformer and dynamic GCN is proposed, which perform the best OA in classifying lithologic

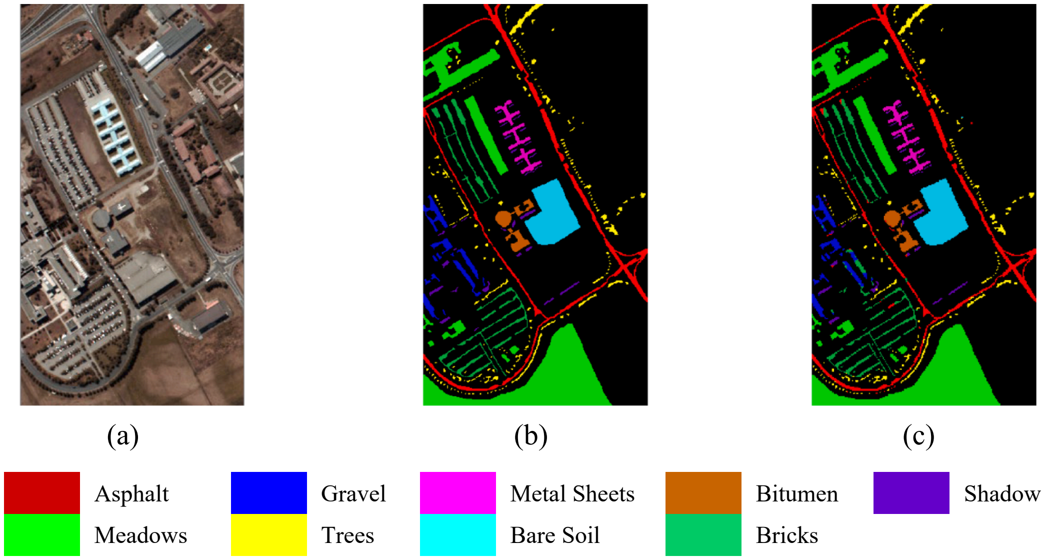


Fig. 19. Visualization of PaviaU dataset. (a) False-color map, (b) Ground truth map, (c) ViT-DGCN classification map.

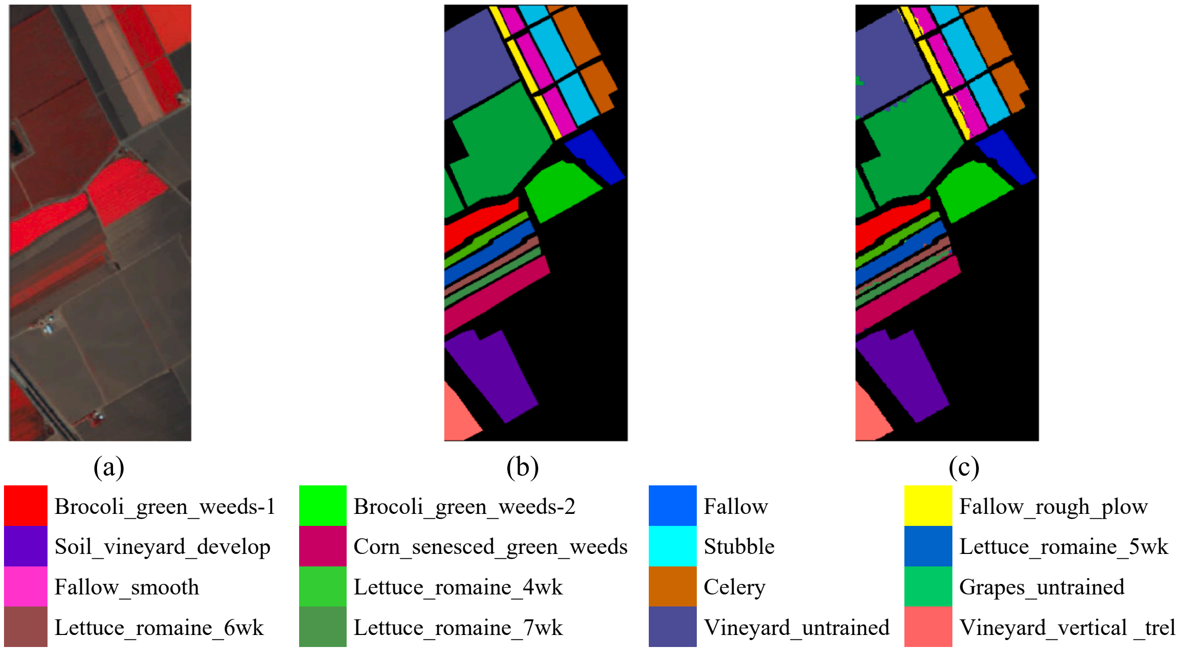


Fig. 20. Visualization of Salinas dataset. (a) False-color map, (b) Ground truth map, (c) ViT-DGCN classification map.

Table 8
Classification accuracies for the PaviaU and Salinas datasets.

Dataset		DFFN	DHCNet	SSRN	FDSSC	Spectral Former	FDGC	ViT-DGCN
PaviaU	OA (%)	92.13	95.71	96.92	97.85	93.78	97.90	**98.52**
	kappa × 100	89.51	95.33	95.90	96.92	91.69	97.06	**98.03**
Salinas	OA (%)	93.61	95.28	95.85	94.34	96.58	97.81	**98.78**
	kappa × 100	92.91	94.75	95.39	93.66	96.19	96.89	**98.64**

units compared to other DL models.

The classification results show excellent classification accuracy and a remarkable recognition effect, providing a successful application example for other similar mineralization areas. The large-scale expansion and application of this technical process will provide field geologists with accurate prospecting targets and technical support for exploring rare metal metallogenic belts in the Himalayas. However, it is challenging and continuous to improve the accuracy of geological mapping. In the future, we will include additional RS geological data, such as visual and infrared multispectral imager data, LiDAR data, SAR data, and other diagnostic features for lithological identification and geological mapping using the developed methods.

CRediT authorship contribution statement

Yanni Dong: Writing – original draft, Software, Methodology, Funding acquisition, Conceptualization. **Zhenzhen Yang:** Writing – original draft, Methodology, Data curation, Conceptualization. **Quanwei Liu:** Writing – review & editing, Visualization. **Renguang Zuo:** Writing – review & editing, Conceptualization. **Ziye Wang:** Writing – review & editing, Resources, Investigation.

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

Data will be made available on request.

Acknowledgments

This work was supported by the National Natural Science Foundation of China under Grants 62222116, 62171417, 42102332 and 42172326.

References

An, Z., Shi, Z., 2014. An improved-SFIM fusion method based on the calibration process. Optik. 125 (14), 3764–3769. https://doi.org/10.1016/j.ijleo.2014.03.005.

Bachri, I., Hakdaoui, M., Raji, M., Teodoro, A.C., Benbouziane, A., 2019. Machine learning algorithms for automatic lithological mapping using remote sensing data: a case study from souk arbaa Sahel, Sidi Ifni inlier, Western anti-atlas. Morocco. ISPRS Int. J. Geoinf. 8 (6), 248. https://doi.org/10.3390/ijgi8060248.

Behnia, P., 2005. Comparison between four methods for data fusion of ETM+ multispectral and pan images. Geo. Spat. Inf. Sci. 8 (2), 98–103. https://doi.org/10.1007/BF02826847.

Bhan, S.K., Krishnanunni, K., 1983. Applications of remote sensing techniques to geology. Proc. Indian Acad. Sci. (engg. Sci.) 6 (4), 297–311. https://doi.org/10.1007/BF02881136.

Brandmeier, M., Chen, Y., 2019. Lithological classification using multi-sensor data and convolutional neural networks. Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci. 42, 55–59. https://doi.org/10.5194/isprs-archives-XLII-2-W16-55-2019.

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

Cao, H., Li, G., Zhang, R., Zhang, R., Zhang, L., Dai, Z., Zhang, Z., Liang, W., Dong, S., Xia, X., 2021. Genesis of the cuonadong tin polymetallic deposit in the tethyan himalaya: evidence from geology, geochronology, fluid inclusions and multiple isotopes. Gondwana Res. 92, 72–101. https://doi.org/10.1016/j.gr.2020.12.020.

Cao, H., Li, G., Zhang, L., Zhang, X., Yu, X., Chen, Y., Lin, B., Pei, Q., Tang, L., Zou, H., 2022. Genesis of himalayan leucogranite and its potentiality of rare metal mineralization. Sediment. Geol. Tethyan Geol. 42 (2), 189–211. https://doi.org/10.19826/j.cnki.1009-3850.2022.04004.

Cooley, T., Anderson, G.P., Felde, G.W., Hoke, M.L., Ratkowski, A.J., Chetwynd, J.H., Gardner, J.A., Adler-Golden, S.M., Matthew, M.W., Berk, A., Bernstein, L.S., Acharya, P.K., Miller, D., Lewis, P., 2002. FLAASH a MODTRAN4-based atmospheric correction algorithm its application and validation. Proc. Int. Geosci. Remote Sens. Symp. 3, 1414–1418. https://doi.org/10.1109/IGARSS.2002.1026134.

Dang, L., Weng, L., Hou, Y., Zuo, X., Yang, L., 2023. Double-branch feature fusion transformer for hyperspectral image classification. Sci. Rep. 13, 272. https://doi.org/10.1038/s41598-023-27472-z.

Deng, T., Sharafat, A., Wie, Y.M., Lee, K.G., Lee, E., Lee, K.H., 2023. A geospatial analysis-based method for railway route selection in marine glaciers: a case study of the sichuan-tibet railway network. Remote Sens. 15 (17), 4175. https://doi.org/10.3390/rs15174175.

Ding, W., Ding, L., 2022. Hyperspectral remote sensing of rock and mineral and its application prospects on the tibetan plateau. Chinese J. Geol. 57 (3), 924–944. https://doi.org/10.12017/dzkx.2022.053.

Dosovitskiy, A., Beyer, A., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., Houlsby, N., 2020. An image is worth 16×16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929. Doi: 10.48550/arXiv.2010.11929.

El-Omairi, M.A., El Garouani, A., 2023. A review on advancements in lithological mapping utilizing machine learning algorithms and remote sensing data. Heliyon. 9 (9), e20168.

Fan, Y., Wang, H., 2020. Application of remote sensing to identify copper–Lead–Zinc deposits in the heiqia area of the West Kunlun Mountains. Chinas. Sci. Rep. 10 (1), 1–13. https://doi.org/10.1038/s41598-020-68464-7.

Feng, J., Rogge, D., Rivard, B., 2018. Comparison of lithological mapping results from airborne hyperspectral VNIR SWIR, LWIR and combined data. Int. J. Appl. Earth Obs. Geoinf. 63, 340–353. https://doi.org/10.1016/j.jag.2017.03.003.

Ge, W., Cheng, Q., Tang, Y., Jing, L., Gao, C., 2018. Lithological classification using sentinel-2A data in the shibanjing ophiolite complex in Inner Mongolia. China. Remote Sens. 10 (4), 638. https://doi.org/10.3390/rs10040638.

Girija, R.R., Mayappan, S., 2019. Mapping of mineral resources and lithological units: a review of remote sensing techniques. Int. J. Image Data Fusion 10 (2), 79–106. https://doi.org/10.1080/19479832.2019.1589585.

Hong, D., Han, Z., Yao, J., Gao, L., Zhang, B., Plaza, A., Chanussot, J., 2021. SpectralFormer: rethinking hyperspectral image classification with transformers. IEEE Trans. Geosci. Remote Sens. 60, 1–15. https://doi.org/10.1109/TGRS.2021.3130716.

Kabolizadeh, M., Rangzan, K., Mousavi, S.S., Azhdari, E., 2022. Applying optimum fusion method to improve lithological mapping of sedimentary rocks using sentinel-2 and ASTER satellite images. Earth Sci. Inform. 15, 1765–1778. https://doi.org/10.1007/s12145-022-00836-1.

Khashaba, S.M.A., El-Shibiny, N.H., Hassan, S.M., Takazawa, E., Khedr, M.Z., 2023. Application of remote sensing data integration in detecting mineralized granitic zones: a case study of the gabal al-ijlah Al-Hamra, central Eastern Desert. Egypt. J. African Earth Sci. 200, 104855 https://doi.org/10.1016/j.jafrearsci.2023.104855.

Kim, I.S., Latif, K., Kim, J., Sharafat, A., Lee, D.E., Seo, J., 2022. Vision-based activity classification of excavators by bidirectional LSTM. Appl. Sci. 13 (1), 272. https://doi.org/10.3390/app13010272.

Latif, K., Sharafat, A., Seo, J., 2023. Digital twin-driven framework for TBM performance prediction, visualization, and monitoring through machine learning. Appl. Sci. 13 (20), 11435. https://doi.org/10.3390/app132011435.

Lee, S., Bae, J.Y., Sharafat, A., Seo, J., 2024. Waste lime earthwork management using drone and BIM technology for construction projects: the case study of urban development project. KSCE J. Civ. Eng. 28, 517–531. https://doi.org/10.1007/s12205-023-1245-z.

Lee, C.P., Lim, K.M., Song, Y., Alqahtani, A., 2023. Plant-CNN-ViT: plant classification with ensemble of convolutional neural networks and vision transformer. Plants. 12 (14), 2642. https://doi.org/10.3390/plants12142642.

Li, C., Liu, L., Wang, J., Zhao, C., Wang, R., 2004. Comparison of two methods of the fusion of remote sensing images with fidelity of spectral information. IEEE Int. Geosci. Remote Sens. Symp. 4, 2561–2564. https://doi.org/10.1109/IGARSS.2004.1369819.

Li, G., Zhang, L., Jiao, Y., Xia, X., Dong, S., Fu, J., Liang, W., Zhang, Z., Wu, J., Dong, L., Huang, Y., 2017. First discovery and implications of cuonadong superlarge be-W-sn polymetallic deposit in himalayan metallogenic belt, southern Tibet. Min. Depos. 36 (4), 1003–1008. https://doi.org/10.16111/j.0258-7106.2017.04.014.

Liu, J.G., 2000. Smoothing filter-based intensity modulation: a spectral preserve image fusion technique for improving spatial details. Int. J. Remote Sens. 21 (18), 3461–3472. https://doi.org/10.1080/014311600750037499.

Liu, Q., Dong, Y., Zhang, Y., Luo, H., 2022. A fast dynamic graph convolutional network and CNN parallel network for hyperspectral image classification. IEEE Trans. Geosci. Remote Sens. 60, 1–15. https://doi.org/10.1109/TGRS.2022.3179419.

Liu, L., Feng, J., Rivard, B., Xu, X., Zhou, J., Han, L., Yang, J., Ren, G., 2018. Mapping alteration using imagery from the Tiangong-1 hyperspectral spaceborne system: example for the jintanzi gold province, China. Int. J. Appl. Earth Obs. Geoinf. 64, 275–286. https://doi.org/10.1016/j.jag.2017.03.013.

Liu, Y., Sun, D., Hu, X., Liu, S., Cao, K., 2019. The advanced hyperspectral imager: aboard China's gaoFen-5 satellite. IEEE Geocsc. Rem. Sen. m. 7 (4), 23–32. https://doi.org/10.1109/MGRS.2019.2927687.

Liu, Y., Sangineto, E., Bi, W., Sebe, N., Lepri, B., Nadai, M., 2021. Efficient training of visual transformers with small datasets. Adv. Neural Inf. Process. Syst. 34, 23818–23830. https://doi.org/10.48550/arXiv.2106.03746.

Lyon, R.J.P., 1972. Infrared spectral emittance in geological mapping: airborne spectrometer data from Pisgah crater. California. Science. 175 (4025), 983–986. https://doi.org/10.1126/science.175.4025.983.

Main-Knorn, M., Pflug, B., Louis, J., Debaecker, V., Müller-Wilm, U., Gascon, F., 2017. Sen2Cor for Sentinel-2. Proc. Image Signal Process. Remote Sens. 10427, 37–48. https://doi.org/10.1117/12.2278218.

Manap, H.S., Bekir, T.S., 2022. Data integration for lithological mapping using machine learning algorithms. Earth Sci. Inform. 15, 1841–1859. https://doi.org/10.1007/s12145-022-00826-3.

Marzouki, A., Dridri, A., 2023. Lithological discrimination and structural lineaments extraction using landsat 8 and ASTER data: a case study of tiwit (anti-atlas, Morocco). Environ. Earth Sci. 125 https://doi.org/10.1007/s12665-023-10831-4.

Ousmanou, S., Fozing, E.M., Kwékam, M., Fodoue, Y., Jeatsa, L.D.A., 2023. Application of remote sensing techniques in lithological and mineral exploration: discrimination of granitoids bearing iron and corundum deposits in southeastern Banyo Adamawa Region-Cameroon. Earth Sci. Inform. 16, 259–285. https://doi.org/10.1007/s12145-023-00937-5.

Pal, M., Rasmussen, T., Porwal, A., 2020. Optimized lithological mapping from multispectral and hyperspectral remote sensing images using fused multi-classifiers. Remote Sens. 12, 177. https://doi.org/10.3390/rs12010177.

Pan, T., Zuo, R., Wang, Z., 2023. Geological mapping via convolutional neural network based on remote sensing and geochemical survey data in vegetation coverage areas. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 16, 3485–3494. https://doi.org/10.1109/JSTARS.2023.3260584.

Peighambari, S., Zhang, Y., 2021. Hyperspectral remote sensing in lithological mapping, mineral exploration, and environmental geology: an updated review. J. Appl. Remote Sens. 15 (3), 031501 https://doi.org/10.1117/1.JRS.15.031501.

Qing, Y., Liu, W., Feng, L., Gao, W., 2021. Improved transformer net for hyperspectral image classification. Remote Sens. 13 (11), 2216. https://doi.org/10.3390/rs13112216.

Ren, K., Sun, W., Meng, X., Yang, G., Du, Q., 2020. Fusing China GF-5 hyperspectral data with GF-1, GF-2 and sentinel-2a multispectral data: which methods should be used? Remote Sens. 12, 882. https://doi.org/10.3390/rs12050882.

Shayeganpour, S., Tangestani, M.H., Gorsevski, P.V., 2021. Machine learning and multi-sensor data fusion for mapping lithology: a case study of kowli-kosh area. SW Iran. Adv. Space Res. 68 (10), 3992–4015. https://doi.org/10.1016/j.asr.2021.08.003.

Shirmard, H., Farahbakhsh, E., Muller, D., Chandra, R., 2022a. A review of machine learning in processing remote sensing data for mineral exploration. Remote Sens. Environ. 268, 112750 https://doi.org/10.1016/j.rse.2021.112750.

Shirmard, H., Farahbakhsh, E., Heidari, E., Pour, A.B., Pradhan, B., Müller, D., Chandra, R., 2022b. A comparative study of convolutional neural networks and conventional machine learning models for lithological mapping using remote sensing data. Remote Sens. 14 (4), 819. https://doi.org/10.3390/rs14040819.

Song, W., Li, S., Fang, L., Lu, T., 2018. Hyperspectral image classification with deep feature fusion network. IEEE Trans. Geosci. Remote Sens. 56 (6), 3173–3184. https://doi.org/10.1109/TGRS.2018.2794326.

Kipf, T.N., Welling, M., 2016. Semi-supervised classification with graph convolutional networks. arXiv preprint arXiv:1609.02907. Doi: 10.48550/arXiv.1609.02907.

Sun, W., Ren, K., Yang, G., Meng, X., Liu Y., 2019. Investigating GF-5 hyperspectral and GF-1 multispectral data fusion methods for multitemporal change Analysis. 2019 10th International Workshop on the Analysis of Multitemporal Remote Sensing Images (MultiTemp), Shanghai, China, 1-4. Doi: 10.1109/Multi-Temp.2019.8866908.

Sun, W., Ren, K., Meng, X., Xiao, C., Yang, G., Peng, J., 2022a. A band divide-and-conquer multispectral and hyperspectral image fusion method. IEEE Trans. Geosci. Remote Sens. 60, 1–13. https://doi.org/10.1109/TGRS.2020.3046321.

Sun, L., Zhao, G., Zheng, Y., Wu, Z., 2022b. Spectral-spatial feature tokenization transformer for hyperspectral image classification. IEEE Trans. Geosci. Remote Sens. 60, 1–14. https://doi.org/10.1109/TGRS.2022.3144158.

Tong, Q., Zhang, B., Zhang, L., 2016. Current progress of hyperspectral remote sensing in China. J. Remote Sens. 20 (5), 689–707. https://doi.org/10.11834/jrs.20166264.

Vivone, G., 2023. Multispectral and hyperspectral image fusion in remote sensing: a survey. Inform. Fusion. 89, 405–417. https://doi.org/10.1016/j.inffus.2022.08.032.

Wan, Y., Fan, Y., Jin, M., 2021. Application of hyperspectral remote sensing for supplementary investigation of polymetallic deposits in huaniushan ore region, northwestern China. Sci. Rep. 11 (1), 1–12. https://doi.org/10.1038/s41598-020-79864-0.

Wang, F., Tax, D.M.J., 2016. Survey on the attention based RNN model and its applications in computer vision. arXiv preprint arXiv:1601.06823. Doi: 10.48550/arXiv.1601.06823.

Wang, W., Dou, S., Jiang, Z., Sun, L., 2018. A fast dense spectral–spatial convolution network framework for hyperspectral images classification. Remote Sens. 10 (7), 1068. https://doi.org/10.3390/rs10071068.

Wang, S., Huang, X., Han, W., Li, J., Zhang, X., Wang, L., 2023a. Lithological mapping of geological remote sensing via adversarial semi-supervised segmentation network. Int. J. Appl. Earth Obs. Geoinf. 125, 103536 https://doi.org/10.1016/j.jag.2023.103536.

Wang, Y., Li, G., Liang, W., Zhang, Z., 2022b. The chemical characteristics and metallogenic mechanism of beryl from cuonadong sn-W-be rare polymetallic deposit in southern Tibet. China. Minerals 12 (5), 497. https://doi.org/10.3390/min12050497.

Wang, X., Tan, K., Du, P., Pan, C., Ding, J., 2022a. A unified multiscale learning framework for hyperspectral image classification. IEEE Trans. Geosci. Remote Sens. 60, 1–19. https://doi.org/10.1109/TGRS.2022.3147198.

Wang, X., Tan, K., Du, P., Han, B., Ding, J., 2023b. A capsule-vectored neural network for hyperspectral image classification. Knowl-Based Syst. 268, 110482 https://doi.org/10.1016/j.knosys.2023.110482.

Wang, Z., Zuo, R., Dong, Y., 2020. Mapping of himalaya leucogranites based on ASTER and Sentinel-2A datasets using a hybrid method of metric learning and random forest. IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens. 13, 1925–1936. https://doi.org/10.1109/JSTARS.2020.2989509.

Wang, Z., Zuo, R., Liu, H., 2021a. Lithological mapping based on fully convolutional network and multi-source geological data. Remote Sens. 13 (23), 4860. https://doi.org/10.3390/rs13234860.

Wang, Z., Zuo, R., Jing, L., 2021b. Fusion of geochemical and remote-sensing data for lithological mapping using random forest metric learning. Math. Geosci. 53, 1125–1145. https://doi.org/10.1007/s11004-020-09897-8.

Wang, Z., Zuo, R., 2022. Mineral prospectivity mapping using a joint singularity based weighting method and long short-term memory network. Comput. Geosci. 158, 104974 https://doi.org/10.1016/j.cageo.2021.104974.

Xi, Y., Taha, A.M.M., Hu, A., Liu, X., 2022. Accuracy comparison of various remote sensing data in lithological classification based on random forest algorithm. Geocarto Int. 128, 109–120. https://doi.org/10.1080/10106049.2022.2088859.

Xiong, Y., Zuo, R., Luo, Z., Wang, X., 2022. A physically constrained variational autoencoder for geochemical pattern recognition. Math. Geosci. 54, 783–806. https://doi.org/10.1007/s11004-021-09979-1.

Yang, X., Cao, W., Lu, Y., Zhou, Y., 2022. Hyperspectral image transformer classification networks. IEEE Trans. Geosci. Remote Sens. 60, 1–15. https://doi.org/10.1109/TGRS.2022.3171551.

Yang, M., Kang, L., Chen, H., Zhou, M., Zhang, J., 2018. Lithological mapping of east tianshan area using integrated data fused by chinese GF-1 PAN and ASTER multi-spectral data. Open Geosci. 10 (1), 532–543. https://doi.org/10.1515/geo-2018-0042.

Ye, B., Tian, S., Ge, J., Sun, Y., 2017. Assessment of WorldView-3 data for lithological map. Remote Sens. 9 (11), 1132. https://doi.org/10.3390/rs9111132.

Ye, B., Tian, S., Cheng, Q., Ge, Y., 2020. Application of lithological mapping based on advanced hyperspectral imager (AHSI) imagery onboard Gaofen-5 (GF-5) satellite. Remote Sens. 12 (23), 3990. https://doi.org/10.3390/rs12233990.

Yokoya, N., Grohnfeldt, C., Chanussot, J., 2017. Hyperspectral and multispectral data fusion: a comparative review of the recent literature. IEEE Trans. Geosci. Remote Sens. 5 (2), 29–56. https://doi.org/10.1109/MGRS.2016.2637824.

Yu, J., Zhang, L., Li, Q., Li, Y., 2021. 3D autoencoder algorithm for lithological mapping using ZY-1 02D hyperspectral imagery: a case study of liuyuan region. J. Appl. Remote Sens. 15 (4), 042610 https://doi.org/10.1117/1.JRS.15.042610.

Zhang, B., Gao, L., Li, J., Hong, D., Zheng, K., 2023. Advances and prospects in hyperspectral and multispectral remote sensing image super-resolution fusion. Acta Geod. Et Cartogr. Sin. 52 (7), 1074–1089. https://doi.org/10.11947/j.AGCS.2023.20220499.

Zhang, X., Li, P., 2014. Lithological mapping from hyperspectral data by improved use of spectral angle mapper. Int. J. Appl. Earth Obs. Geoinf. 31, 95–109. https://doi.org/10.1016/j.jag.2014.03.007.

Zhang, C., Zuo, R., Xiong, Y., 2021. Detection of the multivariate geochemical anomalies associated with mineralization using a deep convolutional neural network and a pixel-pair feature method. Applied Geochem. 130, 104994 https://doi.org/10.1016/j.apgeochem.2021.104994.

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 (2), 847–858. https://doi.org/10.1109/TGRS.2017.2755542.

Zhu, J., Fang, L., Ghamisi, P., 2018. Deformable convolutional neural networks for hyperspectral image classification. IEEE Geosci. Remote s. 15 (8), 1254–1258. https://doi.org/10.1109/LGRS.2018.2830403.

Zou, J., He, W., Zhang, H., 2022. LESSFormer: local-enhanced spectral-spatial transformer for hyperspectral image classification. IEEE Trans. Geosci. Remote Sens. 60, 1–16. https://doi.org/10.1109/TGRS.2022.3196771.

Zuo, R., Xu, Y., 2023. Graph deep learning model for mapping mineral prospectivity. Math. Geosci. 55, 1–21. https://doi.org/10.1007/s11004-022-10015-z.