

Article

A Spectrally Enhanced Multi-Scale CNN for Limited-Sample Lithological Mapping Using Band-Integrated ASTER and Sentinel-2A Imagery

Qiuming Pei ^{1,2,*} , Jiale Shen ¹, Li Zhang ¹, Yifei Zhang ¹, Sergei Krivonogov ^{1,2}, Shiming Wang ^{1,2} and Daren Fang ³

¹ Faculty of Geosciences and Engineering, Southwest Jiaotong University, Chengdu 611756, China

² Sichuan Province Engineering Technology Research Center of Ecological Mitigation of Geohazards in Tibet Plateau Transportation Corridors, Chengdu 611756, China

³ Geoscience Documentation Center, China Geological Survey, Beijing 100083, China

* Correspondence: pqm@swjtu.edu.cn

Highlights

What are the main findings?

- Band-integrated ASTER–Sentinel-2A imagery improves lithological separability by combining complementary spectral information.
- The proposed spectrally enhanced multi-scale CNN outperforms SAM, SVM, and a conventional 3D-CNN under limited-label conditions by jointly strengthening inter-band spectral representation and multi-scale spatial–textural feature extraction.

What are the implications of the main findings?

- Freely available multispectral data, when properly integrated at the band level and combined with a task-oriented CNN architecture, can provide a cost-effective alternative for lithological mapping in arid to semi-arid bedrock-exposed regions where hyperspectral data or dense field samples are unavailable.
- The framework provides a practical basis for improving regional geological mapping and mineral prospecting, while highlighting the need for spatially independent validation and further testing in more complex surface-cover conditions.

Abstract

Lithological mapping with multispectral remote sensing remains challenging when diagnostic spectral information is limited and reliable labeled samples are scarce. This problem is particularly relevant when convolutional neural networks (CNNs) are applied to lithological classification, because limited spectral dimensionality and scarce training samples may hinder the learning of discriminative spatial–spectral features. In this study, we developed a limited-sample lithological mapping framework for the Shibaocheng area of Subei County, Gansu Province, China, using band-integrated ASTER and Sentinel-2A multispectral imagery. ASTER shortwave infrared (SWIR) bands were co-registered and resampled to Sentinel-2A imagery, and then integrated with Sentinel-2A visible and near-infrared (VNIR) and red-edge bands to construct a complementary multispectral dataset. A compact spectrally enhanced multi-scale CNN was designed, incorporating a residual spectral feature enhancement module for inter-band representation learning and a parallel multi-scale hybrid convolution module for capturing spatial–spectral features. Eight lithological units were classified under limited-label conditions using 8158 training samples and 3497 spatially independent validation samples. Experimental results show that the band-integrated ASTER–Sentinel-2A dataset improved classification performance compared with



Academic Editor: Chaowei Yang

Received: 9 May 2026

Revised: 26 June 2026

Accepted: 28 June 2026

Published: 3 July 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and

conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

single-sensor inputs. Using the proposed model, the band-integrated dataset achieved an overall accuracy (OA) of 94.12%, average accuracy (AA) of 94.04%, and Kappa coefficient of 0.932, compared with OA values of 93.14% and 92.40% obtained using ASTER and Sentinel-2A alone, respectively. The positive effect of band-level integration was also observed for spectral angle mapper (SAM), support vector machine (SVM), and 3D-CNN, whose OA values increased to 54.33%, 86.12%, and 92.29%, respectively. The proposed CNN achieved the highest OA among the evaluated methods, outperforming SAM, SVM, and the conventional 3D-CNN. In addition, t-SNE visualization indicated that incorporating spatial texture features produced more compact and better-separated lithological clusters than using spectral features alone. Ablation experiments further demonstrated that the proposed spectral feature enhancement and multi-scale hybrid convolution modules each contributed to improving lithological classification performance. These results demonstrate that integrating freely available multispectral data with a lightweight spectral-spatial CNN provides a practical and cost-effective solution for lithological mapping in bedrock-exposed arid to semi-arid regions, especially where hyperspectral imagery and dense field samples are unavailable.

Keywords: lithological mapping; convolutional neural networks; band integration; spectral-spatial features; Subei County

1. Introduction

Lithological mapping is a fundamental task in the geosciences, as accurate lithological information is essential for reconstructing tectonic evolution and surface processes, guiding mineral and energy exploration, assessing geohazards, and supporting engineering and environmental investigations [1–6]. Traditionally, lithological maps are produced through field surveys, visual interpretation of aerial or satellite images, and expert-based geological synthesis. Although field mapping remains indispensable, it is often labor-intensive, costly, and difficult to implement efficiently in mountainous, desert, or otherwise inaccessible regions. These limitations are particularly pronounced in areas characterized by complex lithological assemblages, strong weathering, sparse field control, or extensive surficial cover.

The rapid development of optical remote sensing has provided new opportunities for lithological mapping over large and geologically complex regions [7–11]. Multispectral and hyperspectral sensors can capture diagnostic absorption and reflectance characteristics related to rock-forming minerals, alteration assemblages, and surficial materials. In parallel, artificial intelligence (AI) has substantially improved the automation and reproducibility of remote-sensing-based lithological interpretation [12–17]. Early AI-assisted lithological mapping mainly relied on conventional machine learning algorithms, such as Support Vector Machine (SVM), Random Forest (RF), and k-nearest neighbors (KNN), which can model nonlinear relationships between spectral responses and lithological classes and have been widely applied in geological remote sensing [18–20].

In recent years, deep learning algorithms, particularly convolutional neural networks (CNNs), have become increasingly important for lithological classification and mapping. Compared with conventional machine learning methods that typically depend on hand-crafted features or shallow spectral descriptors, CNNs can automatically learn hierarchical representations from image data and jointly exploit spectral, spatial, and textural information. This capability is especially valuable for discriminating lithologies with similar spectral shapes but different spatial patterns, or for mapping lithological boundaries in structurally complex regions [21–23]. For example, Zheng et al. (2025) applied a CNN to

ASTER and Landsat 8 OLI multispectral imagery for granitoid mapping and demonstrated the effectiveness of spectral–spatial feature learning for lithological discrimination [23]. Similarly, Wu and Liu (2025) developed an interpretable dual-channel CNN based on multi-source remote-sensing data and achieved improved lithological identification performance in complex geological environments [24]. These studies demonstrate that CNN-based methods can effectively exploit multispectral and multisource remote-sensing data for lithological mapping. More recently, geologically informed feature representations and deep-learning strategies have been introduced to further enhance classification accuracy and model interpretability [25].

Despite these advances, CNN-based lithological mapping using multispectral remote sensing imagery still faces three major challenges. First, although hyperspectral imagery can provide abundant diagnostic spectral information and has achieved excellent performance in lithological mapping [26–28], its acquisition is often costly, spatially limited, and unavailable for many geological regions. Consequently, there remains a need to develop accurate lithological mapping approaches based on freely available multispectral datasets. Second, freely available multispectral datasets, such as ASTER and Sentinel-2, are more accessible but individually contain only a limited number of spectral bands. This limited spectral dimensionality can reduce the separability of lithologies with subtle mineralogical or weathering-related differences [29]. Third, CNNs usually require a sufficiently large number of labeled samples for robust training. In geological mapping, however, high-quality labels are commonly scarce because field verification is costly, lithological classes are often imbalanced, and reliable boundaries may be difficult to define from existing geological maps [30–32]. Under limited-sample conditions, standard CNNs may overfit the training data and exhibit poor generalization.

To address these challenges, this study develops a limited-sample lithological mapping framework that combines multi-source multispectral band integration with a compact task-oriented CNN architecture. At the data level, ASTER shortwave infrared (SWIR) bands are co-registered, resampled, and concatenated with Sentinel-2A visible and near-infrared (VNIR) bands to generate a synergistic band-integrated dataset. This integration strategy aims to combine the diagnostic SWIR sensitivity of ASTER to mineralogical variations with the finer spatial detail and VNIR spectral sampling of Sentinel-2A. At the model level, a spectrally enhanced multi-scale CNN is proposed to improve feature extraction from band-integrated multispectral imagery while avoiding excessive model complexity. The network integrates a residual spectral feature enhancement (SFE) module for inter-band feature learning and a parallel multi-scale hybrid convolution module for capturing spatial–spectral patterns at different neighborhood scales.

The main contributions of this study are as follows:

- (1) A cost-effective ASTER–Sentinel-2A band-integration strategy is designed for lithological mapping using freely available multispectral data.
- (2) A compact spectrally enhanced multi-scale CNN is developed to improve lithological classification under limited-label conditions.
- (3) The effects of multi-source data integration and spatial texture information on lithological separability are analyzed using classification experiments and t-SNE visualization.
- (4) The proposed framework is validated in the Shibaocheng area, a geologically complex region along the northern margin of the Qinghai–Tibet Plateau.

2. Geological Setting

The study area is located in and around the Shibaocheng area of Subei County, Gansu Province, northwestern China, on the northern margin of the Qinghai–Tibet Plateau (Figure 1a). Tectonically, it lies near the junction of the Tarim Basin, Dunhuang Terrane,

Qilian Terrane, and Alxa Terrane, and forms part of the tectonic transition zone between the Central Asian Orogenic System and the Tethyan Orogenic System. This region has experienced a prolonged and complex geological evolution, including Neoproterozoic continental breakup, Early Paleozoic subduction–accretion and collision, and Mesozoic–Cenozoic intra-continental deformation [33–37]. As a result, it is characterized by a complicated structural framework, diverse lithological assemblages, and well-developed fault systems.

The Altyn Tagh Fault (ATF) is the dominant tectonic structure in the region (Figure 1b). It is a lithospheric-scale sinistral strike-slip fault that extends for approximately 1600 km along the northern margin of the Tibetan Plateau and records the far-field effects of the India–Eurasia collision [38–40]. In the study area, the ATF separates the Dunhuang Terrane to the north from the western Qilian Terrane to the south. This tectonic configuration has produced contrasting lithological associations on both sides of the fault and has contributed to the spatial juxtaposition of metamorphic, sedimentary, and intrusive rock units. Such geological complexity makes the area suitable for evaluating the performance of remote-sensing-based lithological classification methods.

North of the ATF, the Dunhuang Terrane is largely covered by Tertiary and Quaternary sediments and desert deposits. The exposed basement rocks are dominated by Precambrian metamorphic supracrustal sequences of the Dunhuang Group, which generally trend northeast and consist mainly of quartzite, schist, marble, paragneiss, and other metasedimentary rocks [41]. South of the ATF, the western Qilian Terrane is likewise extensively covered by Cenozoic sediments and desert deposits. Its most widely exposed units include a Proterozoic crystalline basement dominated by granitic gneiss, marble, amphibolite, and minor granulite [34], overlain by Early Paleozoic strata composed of marine volcanic rocks interbedded with carbonate and clastic sediments, as well as Mesozoic sequences dominated by clastic sedimentary rocks [42]. In addition, igneous rocks related to multiple intrusive episodes, especially granitoids, are widely exposed on both sides of the ATF [43]. Subordinate intermediate and mafic rocks also occur within the Qilian Terrane.

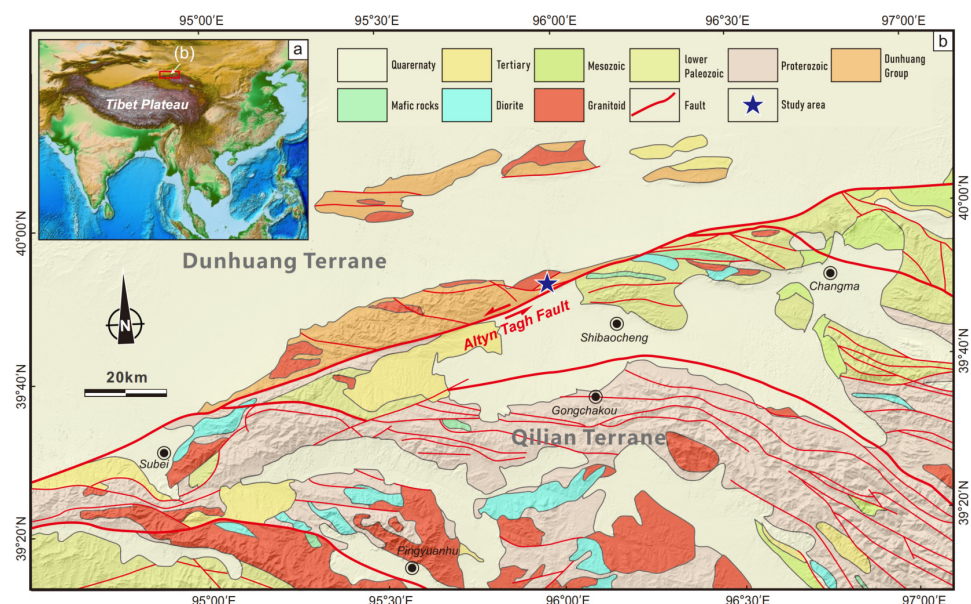


Figure 1. (a) Location of the study area [44]. (b) Simplified geological map of the study area within the Dunhuang Terrane and adjacent tectonic units [34,42].

3. Datasets and Pre-Processing

3.1. Data Sources

This study utilized one ASTER Level-1T scene and one Sentinel-2A Level-1C scene covering the Shibaocheng area, Subei County, Gansu Province, China. The ASTER and Sentinel-2A datasets were acquired on 15 August 2003 and 27 April 2016, respectively. The datasets were obtained from publicly available archives, including the USGS EarthExplorer platform and the Copernicus Data Space Ecosystem. Detailed information on the datasets is summarized in Table 1. ASTER and Sentinel-2A were selected because they provide complementary spectral information for lithological discrimination. ASTER offers six SWIR bands (Bands 4–9; 1.60–2.43 μm), which are highly sensitive to hydroxyl-bearing minerals, carbonates, and hydrothermal alteration minerals commonly used in geological mapping [13,45]. Sentinel-2A provides eight VNIR and red-edge bands (Bands 2–8A) with spatial resolutions of 10–20 m, enabling improved characterization of surface reflectance variability and spatial texture [46,47]. The integration of ASTER mineral-sensitive SWIR information with the higher spatial resolution of Sentinel-2A is therefore particularly suitable for lithological mapping in arid and semi-arid environments. Compared with Landsat 8/9 OLI imagery, which contains only two SWIR bands and fewer VNIR channels, the ASTER–Sentinel-2A combination provides enhanced spectral discrimination capability for lithological mapping. The complementary characteristics of the two sensors facilitate the identification of subtle mineralogical variations and lithological boundaries.

Table 1. Summary of remote sensing datasets used in this study.

Sensor	Product Level	Acquisition Date	Scene ID/Product ID	Spatial Resolution	Spectral Bands Used	NDVI
ASTER	Level-1T	15 August 2003	AST_L1T_00308152003043648_20150430122756_13274.hdf	30 m (SWIR)	Bands 4–9	0.025
Sentinel-2A	Level-1C *	27 April 2016	S2A_MSIL1C_20160427T043712_N0500_R133_T47TKE_20231012T140656.SAFE	10–20 m	Bands 2–8A	0.011

* The Sentinel-2A Level-1C image was atmospherically corrected to surface reflectance during preprocessing.

Although the two datasets were acquired at different times, the study area is characterized by sparse vegetation cover and extensive bedrock exposure. To evaluate the potential influence of temporal differences, the normalized difference vegetation index (NDVI) was calculated for both datasets. As shown in Table 1, the mean NDVI values of the ASTER and Sentinel-2A images were 0.025 and 0.011, respectively. The small difference between these values indicates consistently low vegetation cover and comparable surface conditions, suggesting that temporal differences are unlikely to have a significant impact on lithological mapping results.

3.2. Satellite Data Preprocessing

Prior to band integration, cloud-, cloud-shadow-, and snow-contaminated pixels were removed to minimize their potential influence on subsequent analyses. These interference pixels were identified using empirically derived normalized difference indices and band-based masking techniques implemented in ENVI. Specifically, for ASTER imagery, the Band 8/Band 9 ratio was employed to effectively identify and remove shadowed areas. For Sentinel-2 imagery, the cloud probability layer (MSK_CLDPRB) provided with the downloaded products was used to detect and exclude cloud-contaminated pixels. This preprocessing step improved the consistency and reliability of the multi-source remote sensing data prior to feature extraction and lithological classification.

The ASTER–Sentinel-2A band-integration workflow was implemented in ENVI and comprised four main steps: (1) reprojecting the Sentinel-2A imagery to the ASTER coordinate reference system; (2) resampling the ASTER SWIR bands to the spatial grid of the reprojected Sentinel-2A imagery; (3) co-registering the Sentinel-2A imagery to the ASTER base image using ground control points and image-to-image registration; and (4) stacking the selected Sentinel-2A VNIR bands and ASTER SWIR bands into a unified multispectral reflectance cube.

Specifically, ASTER SWIR Bands 4–9 were stacked with Sentinel-2A Bands 2–8A after geometric co-registration and resampling (Figure 2). Because Sentinel-2A bands have different native spatial resolutions, all selected bands were resampled to a common spatial resolution before band integration. The final band-integrated dataset preserved the diagnostic SWIR information of ASTER while incorporating the finer VNIR spatial and spectral information of Sentinel-2A. This band-integration strategy increased the dimensionality and complementarity of the input features and provided a more informative basis for limited-sample lithological classification.

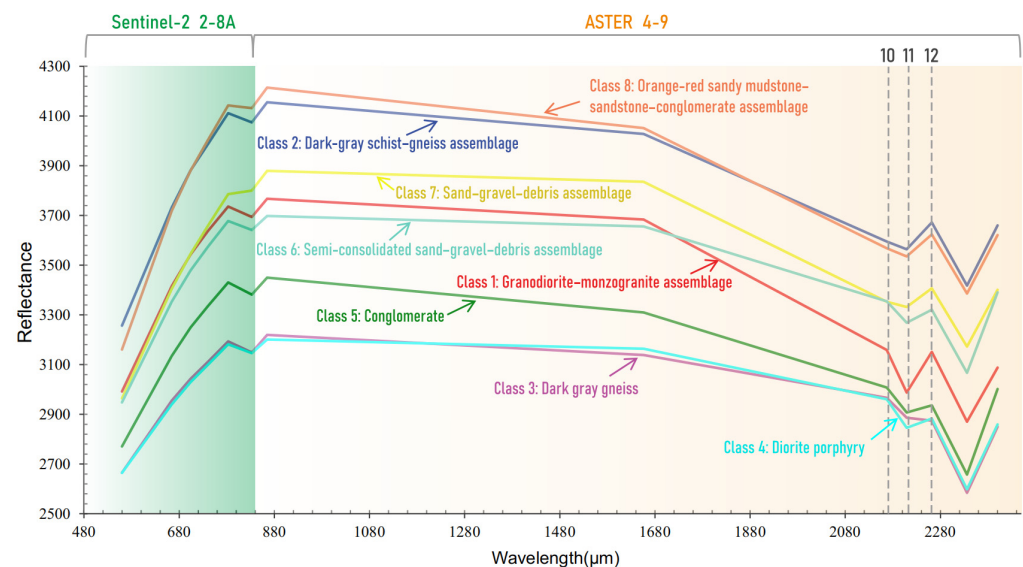


Figure 2. Spectral signatures of the eight lithological classes in the band-integrated ASTER and Sentinel-2A imagery.

3.3. Preparation of Training Data

Training data were prepared by integrating an existing geological map, remote sensing imagery, and expert geological knowledge. A 1:250,000 geological map was first used as the initial reference for delineating lithological boundaries, which were subsequently refined using ASTER false-color composite imagery based on spectral characteristics, surface textures, geomorphological features, and lithological contact relationships (Figure 3). Additional geological information and field observations were also consulted to improve label reliability.

It should be noted that the existing geological map was used only as an auxiliary source for generating reference labels rather than as the final lithological mapping product. The objective of this study is to develop an automated lithological mapping framework for lithological classification under limited-label conditions. Therefore, the proposed framework is intended to reduce the reliance on labor-intensive manual geological mapping while improving the efficiency and transferability of lithological interpretation.

In this study, the training and testing datasets were derived from spatially non-overlapping labeled regions, ensuring spatial independence between the two datasets

and minimizing potential overestimation of model performance caused by spatial auto-correlation. To construct the training and testing datasets, a stratified random sampling strategy was adopted. Lithological classes were used as the stratification variable, and samples were allocated proportionally according to the number of labeled pixels in each class while ensuring sufficient representation of every class in both subsets.

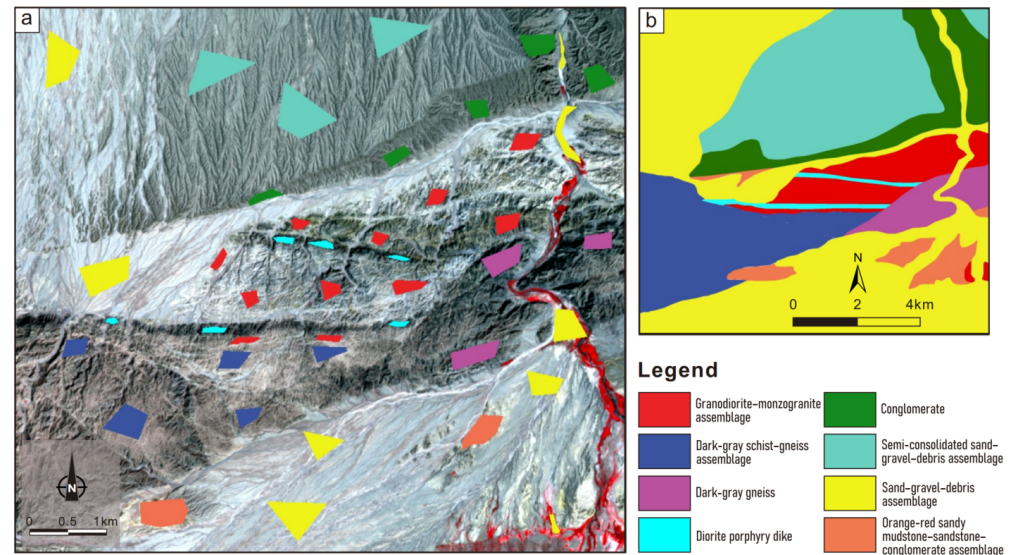


Figure 3. (a) Spatial distribution of collected samples. (b) Geologic map of the study area illustrating lithological units, with boundaries refined from an initial 1:250,000 scale geological map.

4. Satellite Image Processing and Deep Learning Methods

4.1. Conventional Band-Ratio Analysis

Band-ratio techniques were applied to the ASTER and Sentinel-2A imagery to magnify subtle spectral differences between target lithologies while suppressing topographic illumination effects. The selection of specific band combinations was strictly based on the diagnostic physical and chemical absorption features of the predominant rock-forming minerals in the study area. For instance, thermal and shortwave infrared (SWIR) bands are particularly sensitive to the Si-O vibrational features of silicate minerals and the absorption features of carbonates, which are well-documented in the literature for effective lithological discrimination [48,49].

To systematically extract these features, specific mathematical combinations were formulated based on the region's overall mineralogical characteristics. The ratio equations $(\text{Band } 10 + \text{Band } 12) / \text{Band } 11$ and $\text{Band } 11 / \text{Band } 12$ were selected and computed. These calculations were executed in ENVI 5.6 software, serving as both a traditional benchmark for feature enhancement and a baseline comparative method for the subsequent deep learning-based lithological mapping.

4.2. Spectral-Spatial Classification Framework

In this study, a compact CNN was developed to perform lithological classification using the band-integrated dataset. Unlike conventional pixel-based approaches that rely solely on spectral information, the proposed network takes image patches as input and performs lithological classification by jointly exploiting spectral signatures and neighborhood spatial information. Considering the limited availability of labeled geological samples, the network was designed to achieve a balance between feature representation capability, computational efficiency, and model complexity.

The proposed CNN consists of two key feature-extraction modules: (1) an SFE module for improving spectral discrimination, and (2) a Multi-Scale Hybrid Convolution module for extracting spatial-spectral features at different receptive-field scales. These two modules are integrated in sequence and followed by a common fully connected classification head. Together, they progressively learn low-level spectral information, multi-scale spatial features, and high-level semantic representations. The detailed implementation of each component is described in the following subsections.

4.2.1. SFE Module

Multispectral remote sensing imagery contains diagnostic spectral information related to rock-forming minerals and surface materials, making spectral feature learning important for lithological mapping [50]. However, subtle inter-class spectral differences and high intra-class variability may limit feature extraction, while deeper networks can increase the risk of information loss and overfitting under limited-label conditions [51]. Therefore, an SFE strategy was introduced to improve spectral discrimination and preserve informative low-level features.

To address this issue, an SFE module was implemented (Figure 4), inspired by residual learning strategies previously applied in remote sensing image classification [52,53]. The purpose of this module is to enhance spectral discrimination while preserving subtle spectral differences that are important for distinguishing lithological units with similar appearances.

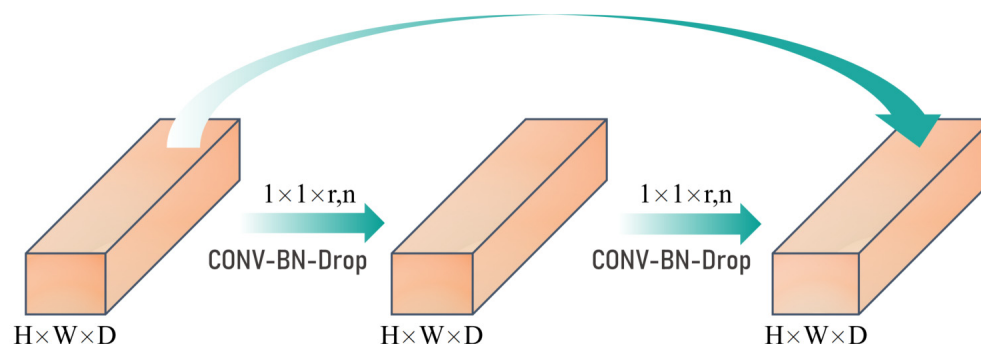


Figure 4. Spectral residual shortcut connections for fusing shallow spectral features with deep learned representations.

Each SFE block consists of two consecutive convolutional layers, followed by batch normalization and a rectified linear unit (ReLU) activation function. Batch normalization was employed to stabilize network training and accelerate convergence, whereas ReLU was used to improve nonlinear feature representation. Residual shortcut connections were subsequently introduced to combine shallow spectral information with deeper learned representations. This design facilitates gradient propagation, mitigates information degradation, and improves feature reuse throughout the network.

Two consecutive SFE blocks were implemented in this study. The first block primarily enhances low-level spectral responses, whereas the second block further refines high-level spectral representations. By progressively integrating shallow and deep spectral information, the network improves the separability of spectrally similar lithological units while maintaining computational efficiency.

4.2.2. Multi-Scale Hybrid Convolution Module

Lithological units exhibit spatial heterogeneity across multiple scales, requiring both local detail extraction and broader contextual representation. However, conventional single-scale convolution with a fixed receptive field may not adequately capture such variability [28,54].

Accordingly, a Multi-Scale Hybrid Convolution module was employed to learn complementary spatial–spectral representations from multiple receptive fields (Figure 5).

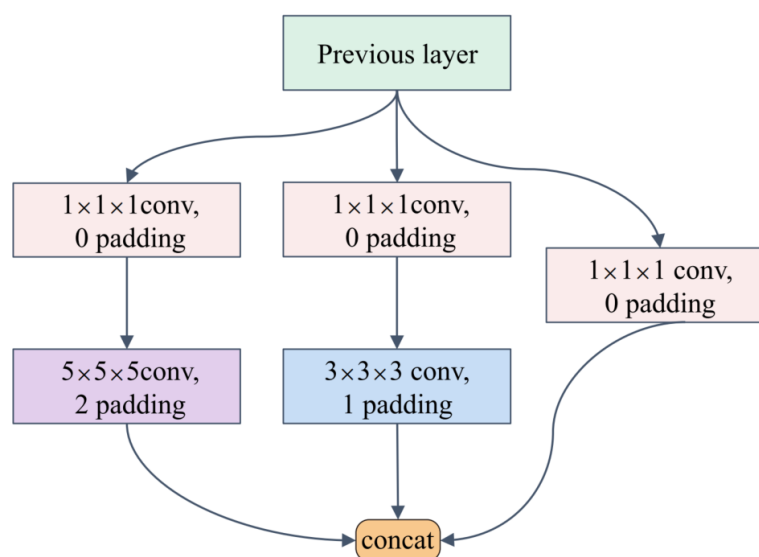


Figure 5. Multi-Scale spatial–spectral information extraction via a three-branch CNN feature extraction flowchart.

The module consists of three parallel convolutional branches, each processing the same input feature map independently. Different convolution kernel sizes are employed to capture lithological characteristics at fine, intermediate, and broad spatial scales. Zero-padding is applied to maintain consistent spatial dimensions across all branches. Each convolution operation is followed by batch normalization and a ReLU activation function. The extracted feature maps are subsequently concatenated along the channel dimension, and dropout regularization is applied to reduce overfitting.

This multi-scale design enables small convolution kernels to focus on local textural details, whereas larger kernels capture broader structural and contextual information. By integrating complementary information from different receptive fields, the module improves the representation of heterogeneous lithological patterns and enhances the discrimination of adjacent lithological units.

4.2.3. Overall Network Architecture

The overall CNN architecture is illustrated in Figure 6. To incorporate neighborhood spatial information, a patch-based input strategy was adopted. Specifically, each center pixel was treated as an individual sample, and a square neighborhood surrounding the center pixel was extracted to form an input patch. The lithological label assigned to the center pixel was then used as the class label for the corresponding patch. For pixels located near image boundaries, out-of-range regions were padded with zeros to maintain a consistent input size [55,56].

Let the input patch be denoted as $S \times S \times D$, where S represents the patch size and D denotes the number of spectral bands in the band-integrated dataset. The input patch first passes through two consecutive SFE blocks to enhance spectral representations and preserve useful low-level spectral information. Subsequently, two multi-scale hybrid convolution blocks are employed to extract joint spatial–spectral features at different receptive-field scales. A max-pooling layer is inserted between the two convolution blocks to reduce feature dimensionality and computational cost while retaining the most discriminative responses. The extracted high-level features are then flattened and fed into two fully connected layers for feature aggregation. Before the final classification stage, dropout regularization is applied to reduce

overfitting and improve model robustness. Finally, a softmax classifier is used to generate the probability distribution of the eight lithological classes.

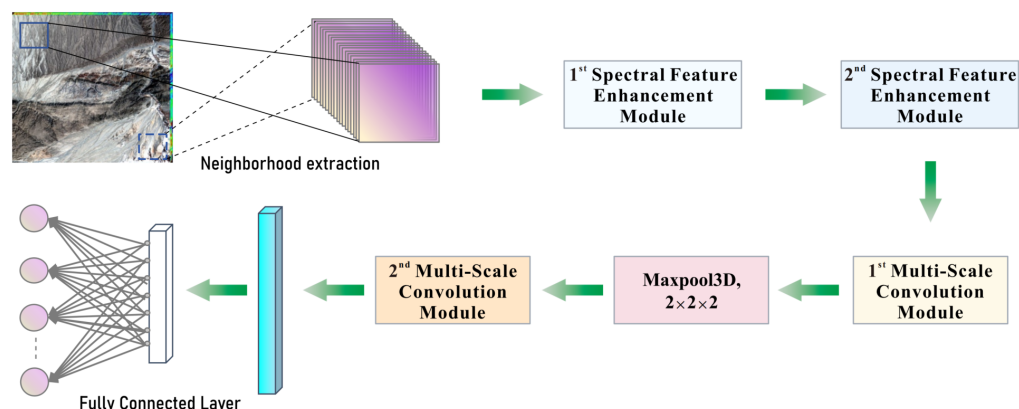


Figure 6. Overall network architecture. Image patches centered on labeled pixels are generated for the study area via neighborhood extraction. Spectral information is enhanced by two consecutive spectral enhancement modules, followed by two multi-scale convolutional modules to capture joint spatial–spectral features. A max-pooling layer is inserted between the two multi-scale convolutional modules to perform downsampling. Finally, two fully connected layers produce the classification output.

Considering the limited availability of labeled samples in lithological mapping, the proposed architecture was intentionally designed to remain relatively shallow in order to avoid over-parameterization. Compared with very deep networks, a lightweight architecture can effectively reduce computational complexity while improving generalization performance under limited-label conditions. Overall, the proposed CNN provides an effective balance between classification accuracy, computational efficiency, and model robustness for multispectral lithological mapping tasks.

4.3. Parameter Settings and Model Training

After cropping to the study area, the band-integrated data cube had a size of $760 \times 813 \times D$, where D denotes the number of spectral bands. A sliding window with a stride of 1 was used to extract image patches, generating 760×813 patches of size $S \times S \times D$ as inputs to the proposed CNN.

The proposed network was implemented, trained, and evaluated using Python and the PyTorch 10.1 framework on a Windows 10 platform. During model training, the cross-entropy loss function was adopted as the optimization objective, and the Adam optimizer was employed to update the network parameters. The training and testing datasets were derived from the labeled regions shown in Figure 3, where a subset of the labeled polygons was used for model training and the remaining polygons were reserved for independent testing. Three key hyperparameters, namely batch size, learning rate, and the number of training epochs, were optimized using a grid-search strategy [23,24,57,58]. The final parameter combination was selected according to classification accuracy and convergence stability (Table 2).

Table 2. Training hyperparameters used for CNN model training.

Hyperparameter	Value
Batch size	64
Learning rate	1×10^{-4}
Number of epochs	50

In addition, the patch size (S) had a significant influence on classification performance because it controls the amount of spatial contextual information available to the model. If S is too small, the model cannot fully exploit neighborhood spatial context; if S is too large, irrelevant background information may be introduced [21,55,56]. This issue is particularly important near lithological boundaries and for small-scale lithologies, such as diorite porphyry dikes, where excessively large patches may contain multiple lithological units and introduce label ambiguity. Therefore, multiple patch sizes were evaluated to balance spatial context representation and lithological purity within each patch. The final patch size was set to $S = 15$, and the corresponding training behavior is presented in Section 5.3.

5. Results and Discussion

5.1. Lithological Units and Sample Statistics

Based on the geological map, ASTER false-color composite imagery, and expert geological knowledge, eight lithological units were interpreted and used as reference labels for supervised classification. Their visual and spatial characteristics are summarized as follows:

- (1) Granodiorite–monzogranite assemblage: This unit was characterized by strongly weathered and fractured surfaces, irregular textures, and silver-white to cyan or blue-green tones in the ASTER false-color composite image.
- (2) Dark-gray schist–gneiss assemblage: This unit was mainly distributed in the north-western part of the study area and generally appeared as brown-toned outcrops.
- (3) Dark-gray gneiss: This unit was characterized by a dark spectral tone, sharp lithological boundaries, and a deep-green appearance in the ASTER false-color composite imagery.
- (4) Diorite porphyry: This unit was predominantly exposed as northwest-trending dikes, exhibiting dark tones and sharp boundaries that make it readily identifiable in the ASTER false-color composite imagery.
- (5) Conglomerate: This unit showed a light dark-green tone and relatively well-defined boundaries.
- (6) Semi-consolidated sand–gravel–debris assemblage: This unit was characterized by pale tones and feather-like surface textures.
- (7) Sand–gravel–debris assemblage: This unit was widely distributed in low-relief areas and exhibited high-reflectance pale-white to light-cyan tones.
- (8) Orange-red sandy mudstone–sandstone–conglomerate assemblage: This unit was sparsely distributed and was characterized by yellow to orange tones, diffuse textures, and poorly defined boundaries.

The labeled dataset exhibited substantial class imbalance across lithological units. For example, the lithological unit with the largest outcrop area, the sand–gravel–debris assemblage, contained 29,448 pixels, whereas the smallest unit, diorite porphyry, contained only 2842 pixels. To mitigate the potential bias caused by this imbalance during model training, the stratified random sampling strategy described in Section 3.3 was applied. The final numbers of total pixels, sampled pixels, training samples, and validation samples are summarized in Table 3.

Table 3. Lithological classes and sample statistics.

Class Label	Class Name	Color	Total Pixels	Number of Sample Pixels	Training Samples	Validation Samples
1	Granodiorite–monzogranite assemblage		23,678	1430	1001	429
2	Dark gray schist–gneiss assemblage		19,302	1263	884	379
3	Dark gray gneiss		9823	997	698	299
4	Diorite porphyry		2842	329	230	99
5	Conglomerate		9223	1101	771	330
6	Semi-consolidated sand–gravel–debris assemblage		9235	2280	1596	684
7	Sand–gravel–debris assemblage		29,448	3279	2295	984
8	Orange-red sandy mudstone–sandstone–conglomerate assemblage		8808	976	683	293

Note: The colors in the “Color” column indicate the display colors assigned to each lithological class in the classification maps.

5.2. Conventional Band-Ratio Results

The application of the designed band-ratio algorithms successfully translated the theoretical spectral differences into visually interpretable images. As illustrated in the spectral profile analysis and the resulting pseudo-color composites (Figure 7), the targeted lithological units exhibit distinct spatial distributions and boundary contrasts. Specifically, the ratio combination of (Band 10 + Band 12)/Band 11 effectively highlights the emissivity and reflectance characteristics of the granodiorite–monzogranite assemblage (Unit 1). In the resulting composite (Figure 7a), this specific rock unit is significantly enhanced, appearing as distinct, high-brightness anomalies. Conversely, the Band 11/Band 12 ratio image successfully isolates the dark-gray gneiss (Unit 3), providing a clear visual demarcation of its spatial extent (Figure 7b). Regarding the Quaternary deposits, represented mainly by Unit 7 (the sand–gravel–debris assemblage) and Unit 8 (the orange-red sandy mudstone–sandstone–conglomerate assemblage), these units generally exhibit relatively high reflectance across most bands, which distinguishes them from the deeper-seated metamorphic and igneous units in the ratio-based visualizations.

Although these conventional band-ratio techniques demonstrate reliability in identifying specific, mineralogically distinct rock units, their overall mapping capabilities remain limited. They are highly dependent on manual visual interpretation and struggle to simultaneously differentiate all eight complex lithological units, especially in areas with high intra-class spectral variability and severe surface weathering. These limitations underscore the necessity for a more robust, automated approach capable of mining deeper, non-linear spatial–spectral features. Consequently, the proposed spectrally enhanced multi-scale CNN was deployed to overcome these traditional bottlenecks, with its training convergence and mapping performance detailed in the following sections.

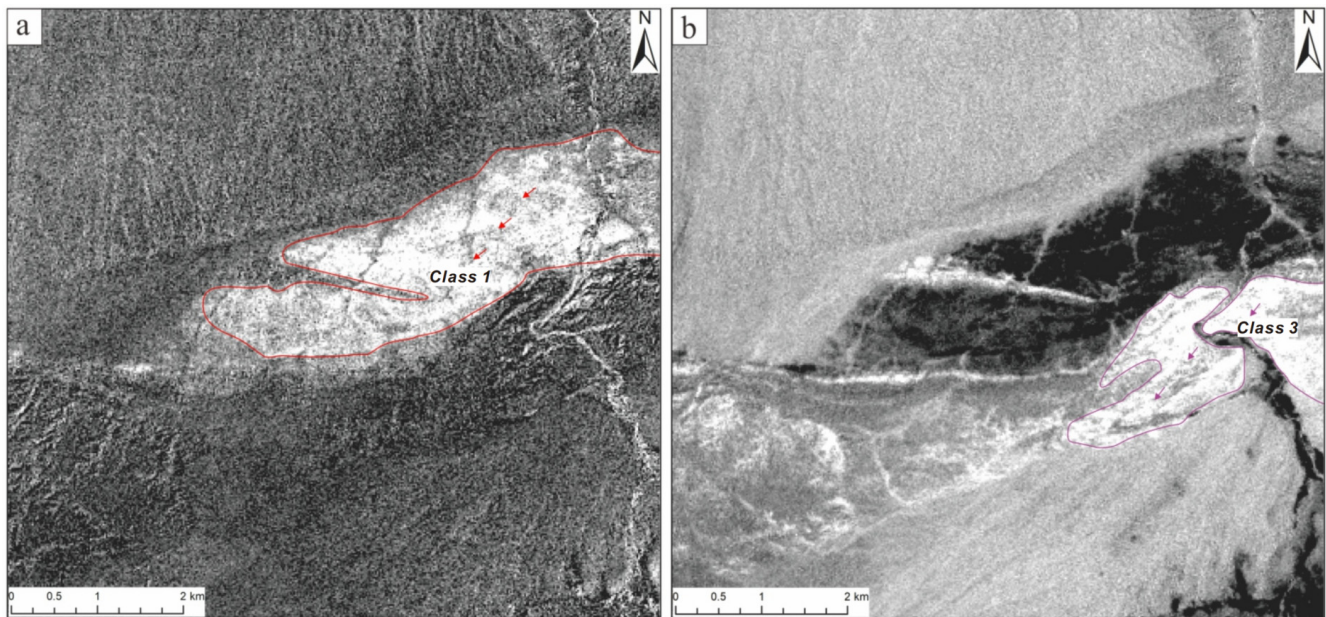


Figure 7. ASTER–Sentinel-2A band-ratio images: (a) (Band 10 + Band 12)/Band 11 and (b) Band 11/Band 12.

5.3. Model Training and Convergence

The model was trained for 50 epochs using the hyperparameters listed in Table 2. Figure 8 presents the variations in loss and classification accuracy for both the training and testing datasets during the training process. During the early training stage, the loss decreased rapidly while the classification accuracy increased steadily. After approximately 20 epochs, both the loss and accuracy curves gradually stabilized, indicating that the model had converged.

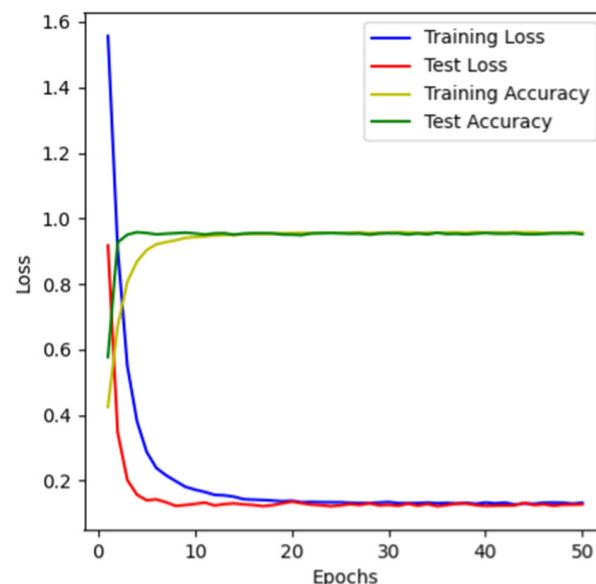


Figure 8. Training and testing performance of the model proposed in this study.

At the end of the training process, the model achieved an accuracy of 95.69% on the training dataset and 95.55% on the testing dataset, demonstrating stable convergence behavior and good generalization under the adopted limited-sample setting. Among the tested patch sizes, $S = 15$ provided a suitable balance between neighborhood contextual information and lithological purity, and was therefore adopted in the final model.

5.4. Comparative Performance of Lithology Classification Among Different Algorithms

To evaluate the contribution of multi-source band integration, all classification methods were tested on the ASTER dataset, the Sentinel-2A dataset, and the band-integrated dataset. Classification performance was assessed using overall accuracy (OA), average accuracy (AA), and the Kappa coefficient [59]. For each algorithm, the parameter configuration yielding the best performance was adopted to ensure a fair comparison. The classification results are summarized in Table 4.

Table 4. Lithological interpretation results from different models using ASTER, Sentinel-2A, and band-integrated remote sensing imagery.

Class	SAM			SVM		
	ASTER	Sentinel-2A	ASTER–Sentinel-2A	ASTER	Sentinel-2A	ASTER–Sentinel-2A
1	84.36%	81.00%	86.57%	89.68%	90.27%	90.33%
2	32.12%	25.97%	34.21%	83.14%	83.44%	84.12%
3	12.11%	12.54%	21.00%	83.52%	80.08%	85.56%
4	52.93%	48.94%	58.63%	86.77%	86.36%	86.97%
5	78.65%	86.10%	81.12%	80.26%	74.54%	80.02%
6	40.20%	34.17%	48.40%	91.45%	92.03%	93.05%
7	79.52%	82.11%	82.73%	91.92%	90.75%	90.88%
8	8.31%	10.60%	12.16%	83.56%	81.31%	84.36%
OA/%	48.27	48.10	54.33	86.31	84.64	86.12
AA/%	49.54	48.69	55.89	86.28	84.84	86.91
Kappa	0.488	0.484	0.547	0.855	0.842	0.857
Class	3D-CNN			This study		
	ASTER	Sentinel-2A	ASTER–Sentinel-2A	ASTER	Sentinel-2A	ASTER–Sentinel-2A
1	86.00%	89.22%	89.43%	99.26%	98.00%	99.53%
2	89.4%	84.37%	96.51%	97.99%	98.1%	98.21%
3	80.27%	80.58%	85.46%	90.54%	92.27%	92.67%
4	89.85%	87.23%	98.86%	96.42%	96.11%	98.92%
5	91.22%	90.91%	99.92%	89.42%	88.61%	89.57%
6	78.4%	80.24%	98.57%	98.97%	99.24%	100.00%
7	86.72%	83.56%	91.91%	93.62%	91.47%	93.64%
8	76.35%	74.23%	78.32%	79.56%	74.74%	79.78%
OA/%	85.23	83.41	92.29	93.14	92.40	94.12
AA/%	84.77	83.76	92.37	93.25	92.31	94.04
Kappa	0.913	0.903	0.919	0.923	0.912	0.932

Overall, classification accuracies obtained from the band-integrated dataset were consistently higher than those from the individual datasets, indicating that the integration of complementary ASTER SWIR and Sentinel-2A VNIR bands enhances lithological discrimination. ASTER provides SWIR information sensitive to mineralogical variations, whereas Sentinel-2A contributes higher-spatial-resolution VNIR information. The complementary characteristics of the two sensors provide richer spectral and spatial information for lithological classification.

Among the evaluated methods, SAM achieved the lowest performance, with an OA of only 54.33% on the band-integrated dataset. Because SAM relies primarily on spectral angle similarity, it is prone to confusion when different lithological units exhibit similar spectral signatures. In contrast, SVM substantially improved classification performance by constructing nonlinear decision boundaries, achieving an OA of 86.12% on the band-integrated dataset. Nevertheless, its pixel-based classification strategy still produced isolated misclassified pixels and fragmented patches.

The 3D-CNN further improved classification performance by jointly exploiting spectral and spatial information, reaching an OA of 92.29% on the band-integrated dataset. Previous

studies have demonstrated that spatial contextual information can significantly enhance remote sensing image classification performance [28,54]. The proposed model achieved the highest classification accuracy (OA = 94.12%, AA = 94.04%, Kappa = 0.932), suggesting that the SFE module and multi-scale spatial feature extraction module further improve the separability of complex lithological units.

The classification maps further support these quantitative results. Extensive misclassification can be observed in the SAM results (Figure 9a). Although SVM improves class discrimination, noticeable salt-and-pepper noise remains (Figure 9b). The 3D-CNN reduces isolated misclassified pixels but still exhibits boundary confusion near narrow dikes and lithological contact zones (Figure 9c). In comparison, the proposed model generates more spatially coherent classification maps, preserving linear geological features such as diorite porphyry dikes and producing clearer lithological boundaries (Figure 9d).

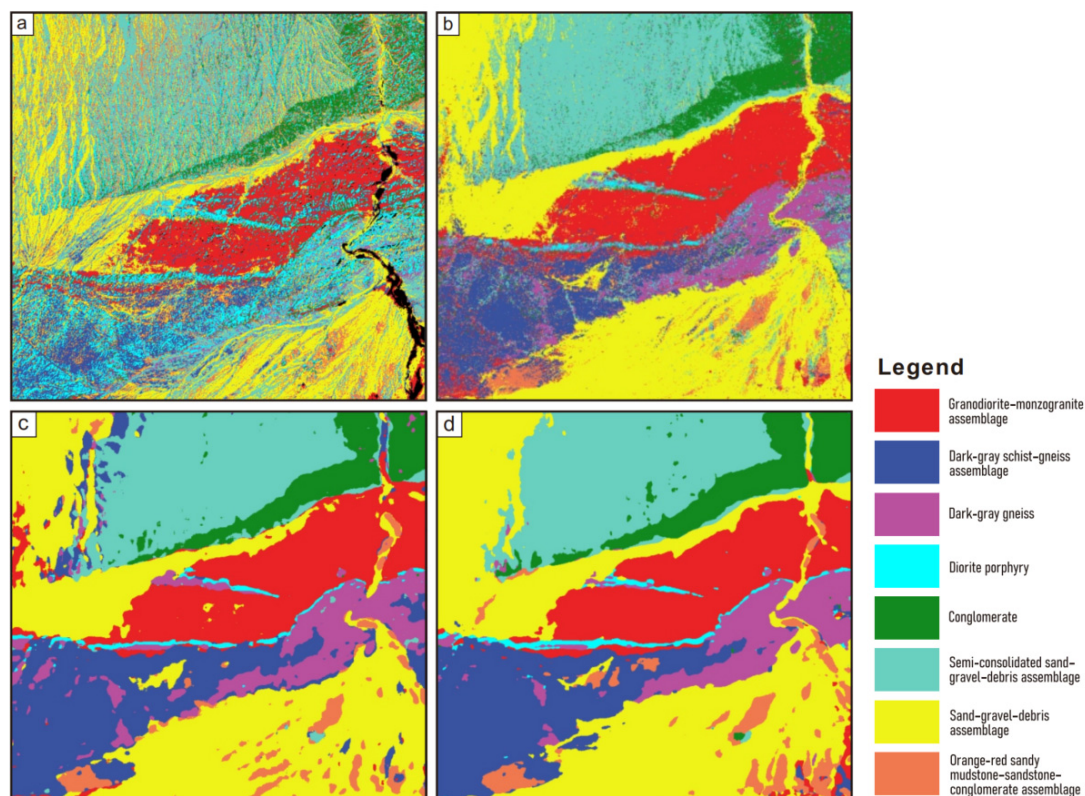


Figure 9. Lithological classification performance of various algorithms using band-integrated ASTER and Sentinel-2A imagery: (a) SAM; (b) SVM; (c) 3D-CNN; and (d) the proposed model.

To improve the reliability of model evaluation, the training and testing samples were selected from spatially independent labeled regions. This strategy minimizes the influence of spatial autocorrelation between neighboring samples and provides a more rigorous assessment of model generalization. The consistently superior performance of the proposed model under this evaluation framework further supports its effectiveness for lithological classification.

5.5. Analysis of the Impact of Multi-Source Band Integration and Spatial Features on Lithological Separability

To further investigate lithological separability, t-distributed stochastic neighbor embedding (t-SNE) was employed to visualize high-dimensional features in a two-dimensional space [57,60]. In remote sensing studies, t-SNE is commonly used to analyze class distributions and feature separability. However, spectral features alone may not fully charac-

terize lithological units that exhibit similar spectral responses but different spatial structures [61,62]. Therefore, gray-level co-occurrence matrix (GLCM)-based texture features were incorporated as descriptors of spatial information [63–65] and combined with spectral features for separability analysis.

Figure 10 presents the t-SNE visualizations generated using spectral features alone and using combined spectral and GLCM texture features. When only spectral features were used, considerable overlap was observed among several lithological classes, and some samples failed to form distinct clusters. This observation indicates that spectral information alone remains insufficient to completely discriminate lithological units with similar mineralogical compositions or comparable weathering characteristics.

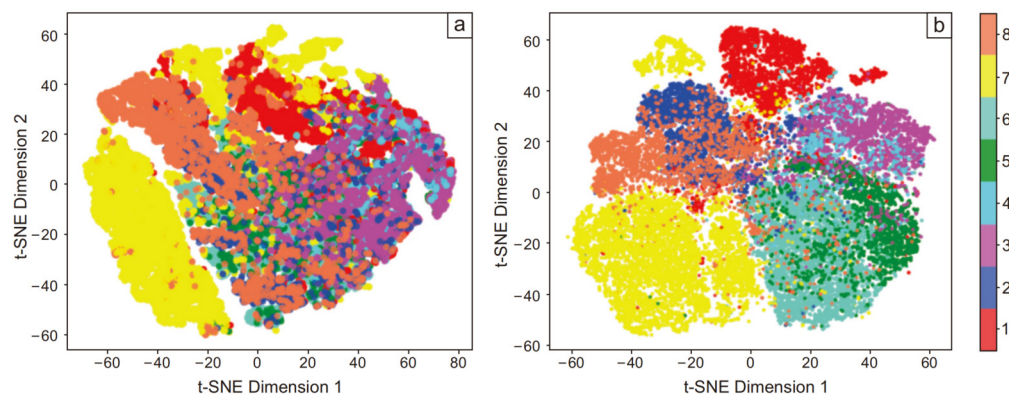


Figure 10. (a) t-SNE visualization of spectral-feature separability; (b) t-SNE visualization of sample separability with GLCM-based spatial features incorporated.

After incorporating GLCM texture features, most lithological classes formed more compact clusters with clearer boundaries. Compared with spectral information alone, texture features characterize local spatial variations, lithological boundaries, and surface structural differences, thereby providing additional discriminative information for lithological identification [63–65]. Consequently, even when different lithologies exhibit similar spectral responses, differences in spatial texture can improve class separability.

These observations are consistent with the classification results presented in Section 5.4. Compared with SAM and SVM, which rely primarily on spectral information, deep learning models that exploit spatial contextual information achieved higher classification accuracies and better spatial continuity [66,67]. Similar findings have been reported in previous studies, where spectral–spatial approaches generally outperform methods based solely on spectral features [28,54]. The improved clustering patterns observed in Figure 10 therefore provide additional evidence that spatial information contributes significantly to lithological discrimination.

Multi-source band integration further enhances lithological separability. The SWIR bands of ASTER are sensitive to hydroxyl-bearing minerals, carbonate minerals, and hydrothermal alteration signatures, whereas Sentinel-2A provides higher spatial resolution and rich VNIR information. The complementary spectral and spatial characteristics of the two sensors enable a more comprehensive representation of lithological features. This interpretation is consistent with the higher classification accuracies obtained from the band-integrated dataset in Table 4, suggesting that the integration of complementary information from both sensors improves the discrimination of adjacent lithological units in complex geological environments.

5.6. Model Advantages, Limitations, and Comparative Analysis with Related Studies

The results demonstrate that the integration of multi-source remote sensing data and deep learning techniques can effectively improve lithological classification perfor-

mance. Compared with conventional spectral classification approaches, the proposed model achieved superior performance in terms of overall accuracy, class discrimination, and spatial consistency. As shown in Table 4, the proposed model achieved an OA of 94.12% on the band-integrated dataset, outperforming SAM, SVM, and 3D-CNN. The improvement is reflected not only in quantitative accuracy metrics but also in the clearer delineation of geological features such as dikes and lithological contact zones. Previous studies have shown that spatial information plays an important role in lithological mapping and that spectral–spatial approaches generally outperform methods based solely on spectral features. Compared with studies using either ASTER or Sentinel data alone, the present study exploits the complementary characteristics of both sensors and jointly models spectral and spatial information within a deep learning framework. Furthermore, unlike many lithological mapping studies based on hyperspectral imagery [26,39], the proposed approach relies on multispectral datasets that are more widely available and suitable for large-area applications.

Despite the encouraging results, several limitations remain. First, the proposed framework was evaluated in a single study area, and lithological assemblages, weathering conditions, and surface cover characteristics may vary considerably in other geological settings. Therefore, its applicability to different regions requires further investigation. Second, the study area is characterized by relatively good rock exposure conditions. Classification performance may decrease in regions with dense vegetation cover, extensive shadows, or complex topography. Finally, although spatially independent training and testing samples were adopted to reduce spatial autocorrelation effects, cross-regional validation has not yet been conducted. Consequently, the generalization capability of the proposed framework remains to be further assessed.

Future work should evaluate the proposed framework across multiple geological regions and investigate the integration of additional data sources, such as DEM, geochemical, and geophysical datasets, to improve lithological mapping performance in more complex environments.

5.7. Ablation Study

To evaluate the individual and joint effects of the two proposed feature-extraction modules, an ablation study was conducted using the band-integrated dataset. The evaluated modules were SFE and Multi-Scale Hybrid Convolution. The fully connected classification head was retained unchanged in all configurations because it served as the common classifier rather than a task-specific module.

Four configurations were evaluated: (1) Baseline CNN, in which both proposed modules were replaced by conventional convolutional blocks; (2) Spec-CNN, which retained only the SFE module; (3) Spa-CNN, which retained only the Multi-Scale Hybrid Convolution module; and (4) Spec-Spa-CNN, which incorporated both the SFE module and the Multi-Scale Hybrid Convolution module. For a fair comparison, all network configurations were trained using identical datasets, training strategies, and hyperparameter settings. The four configurations differed only in the incorporated feature-extraction modules. Classification performance was evaluated using OA, AA, and the Kappa coefficient.

An analysis of the learning curves reveals clear differences in the feature-extraction capabilities of the different network configurations (Figure 11). Specifically, the baseline 3D-CNN exhibits suboptimal convergence and struggles to reach an ideal fitting state even after 50 training epochs. This suggests that relying solely on a conventional 3D-CNN architecture is insufficient for effectively capturing the complex spatial–spectral features of diverse lithological units. However, introducing either the SFE module or the Multi-Scale Hybrid Convolution module results in faster and more stable convergence.

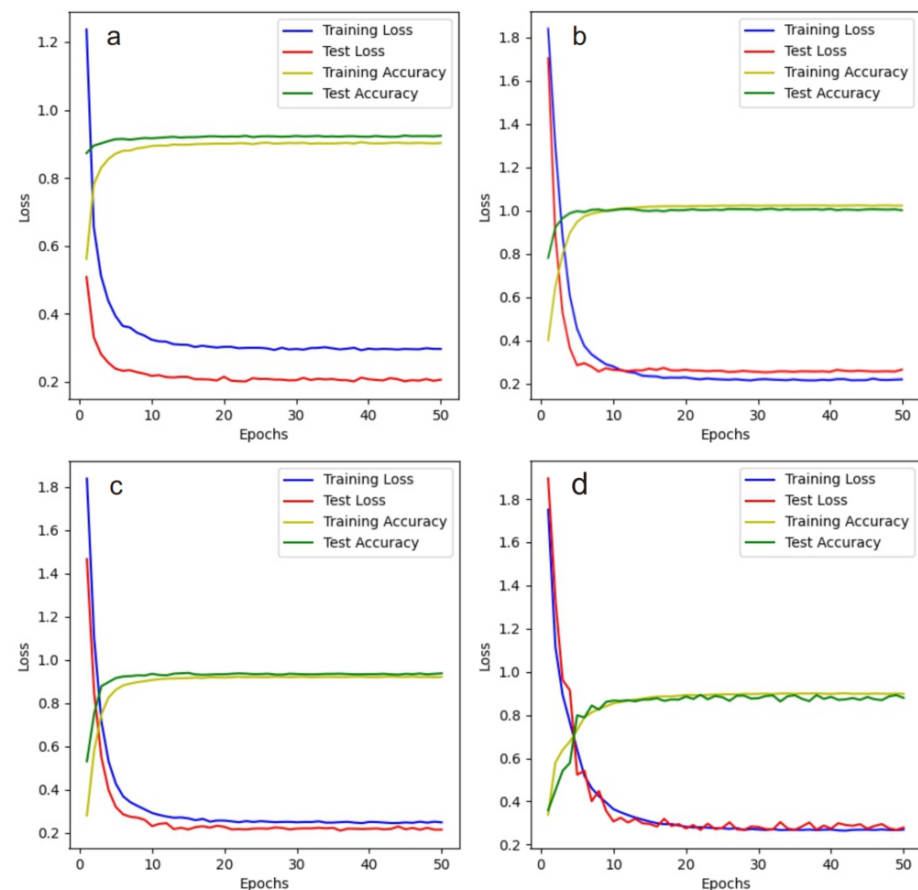


Figure 11. Training and testing convergence curves (loss and accuracy) of different network configurations in the ablation study: (a) the baseline 3D-CNN; (b) the Spec-CNN; (c) the Spa-CNN; and (d) Spec-Spa-CNN.

When both the SFE module and the Multi-Scale Hybrid Convolution module are incorporated simultaneously, the Spec-Spa-CNN reaches a well-fitted state on the training set after approximately 20 epochs. Nevertheless, the loss and accuracy curves on the testing set remain unstable and do not converge satisfactorily even after 50 epochs, suggesting reduced generalization capability. This phenomenon indicates that introducing either feature-extraction module individually effectively improves the overall classification performance, whereas combining both modules increases model complexity and the number of trainable parameters. Under the limited-sample condition of this study area, the increased complexity may weaken the model's generalization ability.

As summarized in Table 5, both proposed modules improved classification performance compared with the baseline 3D-CNN when they were introduced individually. After introducing the SFE module, OA increased from 92.29% to 94.70%, AA increased from 92.37% to 94.52%, and the Kappa coefficient increased from 0.919 to 0.932. This improvement indicates that the residual-based spectral enhancement strategy effectively preserves useful inter-band information during repeated convolution operations and strengthens the discrimination capability of the network for lithological units with similar spectral characteristics. Similarly, incorporating the Multi-Scale Hybrid Convolution module improved OA from 92.29% to 93.68%, AA from 92.37% to 93.14%, and the Kappa coefficient from 0.919 to 0.930. This result suggests that multi-scale feature extraction effectively captures lithological characteristics at different spatial scales and improves the representation of heterogeneous geological patterns, particularly near lithological boundaries and small-scale geological bodies.

Table 5. Performance comparison of different network configurations in the ablation study.

Class	Baseline (3D-CNN)	Spec-CNN	Spa-CNN	Spec-Spa-CNN
1	89.43%	99.41%	98.99%	99.22%
2	96.51%	98.13%	89.43%	94.56%
3	85.46%	92.45%	88.61%	91.45%
4	98.86%	97.94%	96.02%	95.37%
5	99.92%	84.57%	88.95%	82.12%
6	98.57%	100.00%	98.15%	98.56%
7	91.91%	94.35%	96.98%	91.35%
8	78.32%	77.71%	82.43%	76.53%
OA/%	92.29	94.70	93.68	92.64
AA/%	92.37	94.52	93.14	92.56
Kappa	0.919	0.932	0.930	0.928

However, when the SFE module and the Multi-Scale Hybrid Convolution module were incorporated simultaneously, the Spec-Spa-CNN did not achieve further performance improvement over the single-module configurations. Its OA, AA, and Kappa coefficient were 92.64%, 92.56%, and 0.928, respectively, which were lower than those of both Spec-CNN and Spa-CNN. In other words, although the two modules are individually effective, their combination does not necessarily produce an additive performance gain when the available training samples are limited.

The results indicate that the two modules improve classification performance from different perspectives. The SFE module primarily enhances spectral representation learning by preserving subtle spectral differences among lithological units, whereas the Multi-Scale Hybrid Convolution module focuses on extracting complementary spatial contextual information at different receptive-field scales. Therefore, both modules effectively compensate for the limitations of conventional 3D-CNN models in lithological discrimination. However, the more complex Spec-Spa-CNN shows reduced generalization under limited-sample conditions, resulting in lower performance than the single-module configurations.

Figure 12 further illustrates the visual classification results of the ablation experiments. Compared with the baseline model, both Spec-CNN and Spa-CNN produce more spatially coherent classification maps with fewer isolated misclassified pixels. The improvement is particularly evident in areas containing complex lithological boundaries and small-scale lithological units. However, when the SFE module and the Multi-Scale Hybrid Convolution module are used simultaneously, the Spec-Spa-CNN does not show further visual improvement compared with the single-module configurations. Instead, some local classification confusion and fragmented patches remain, which is consistent with the quantitative results in Table 5.

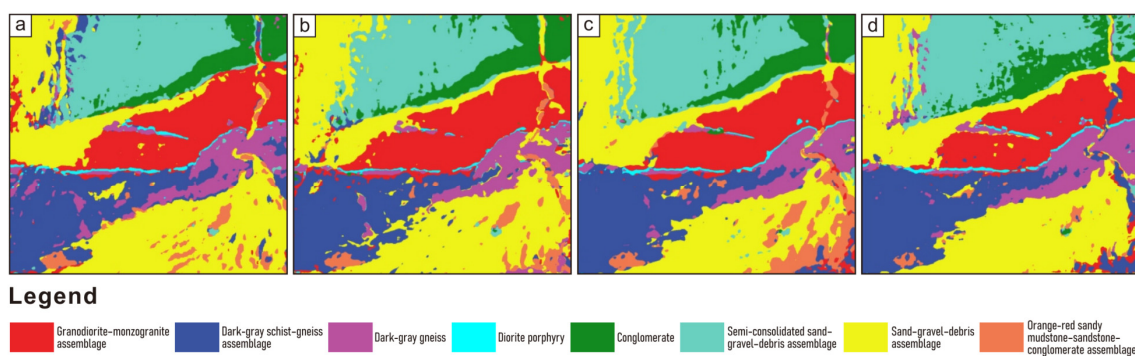


Figure 12. Visual comparison of lithological classification results obtained from different network configurations in the ablation study: (a) baseline 3D-CNN; (b) Spec-CNN; (c) Spa-CNN; and (d) Spec-Spa-CNN.

6. Conclusions

This study developed a spectrally enhanced multi-scale CNN framework for limited-sample lithological mapping using band-integrated ASTER and Sentinel-2A multispectral imagery and evaluated its performance in the Shibaocheng area, Subei County, Gansu Province, China. The results demonstrate that band-level integration of ASTER SWIR bands and Sentinel-2A VNIR bands provides complementary spectral information and improves lithological separability compared with single-sensor datasets. Classification experiments consistently showed that the integrated dataset enhanced mapping performance across different algorithms, indicating the value of combining mineral-sensitive spectral responses with finer spatial information for lithological discrimination.

Compared with SAM, SVM, and the conventional 3D-CNN, the proposed spectrally enhanced multi-scale CNN achieved higher OA, AA, and Kappa coefficient under the experimental setting adopted in this study. These results demonstrate that the proposed framework effectively integrates complementary spectral and spatial information, leading to more robust lithological discrimination under limited-sample conditions. The t-SNE visualization further indicated that the proposed framework generated more compact intra-class feature distributions and clearer inter-class boundaries, demonstrating its enhanced capability for spectral–spatial feature learning and lithological discrimination. The ablation study further confirmed the effectiveness of the residual SFE module in improving inter-band feature representation and the multi-scale convolution strategy in strengthening the extraction of spatial–spectral information at different receptive-field scales, thereby validating the proposed network design.

However, because the framework was validated only in the Shibaocheng area, which is characterized by relatively good bedrock exposure and arid to semi-arid conditions, the present results should be interpreted as demonstrating the potential of the proposed method for limited-sample lithological mapping in similar rocky outcrop environments rather than establishing its general applicability to vegetated or topographically complex regions. Future work should further evaluate the spatial transferability and robustness of the framework through cross-regional validation and testing under more heterogeneous surface conditions. Integrating additional information sources, such as DEM, SAR, geophysical, and geochemical datasets, may further improve lithological mapping performance in areas with stronger vegetation influence and more complex terrain conditions.

Author Contributions: Conceptualization, Q.P. and J.S.; Methodology, Q.P. and J.S.; Software, Q.P., J.S. and L.Z.; Investigation, Q.P., Y.Z., S.W. and D.F.; Data curation, L.Z., Y.Z., S.W. and D.F.; Writing—original draft, Q.P. and J.S.; writing—review and editing, Q.P.; Supervision, S.K.; Funding acquisition, Q.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Sichuan Provincial Science and Technology Program (grant number: 2025YENZH0006), the National Natural Science Foundation of China (grant number: W2431031), and the College Students' Innovative Entrepreneurial Training Plan Program (grant number: 2025168).

Data Availability Statement: The ASTER data can be found at the USGS EarthExplorer [<https://earthexplorer.usgs.gov> (accessed on 6 May 2026)], and the Sentinel-2A data can be found at the Copernicus Data Space Ecosystem [<https://dataspace.copernicus.eu> (accessed on 6 May 2026)]. The datasets generated and analyzed during the current study are available from the corresponding author upon reasonable request.

Acknowledgments: The authors are grateful to the editor and four anonymous reviewers for their constructive comments that significantly enhanced this paper.

Conflicts of Interest: The authors declare no conflicts of interest.

References

- Sharpe, T. The birth of the geological map. *Science* **2015**, *347*, 230–232. [\[CrossRef\]](#)
- Leverington, D.W. Discrimination of sedimentary lithologies using Hyperion and Landsat Thematic Mapper data: A case study at Melville Island, Canadian High Arctic. *Int. J. Remote Sens.* **2010**, *31*, 233–260. [\[CrossRef\]](#)
- Pei, Q.; Shen, J.; Wang, S.; Fang, D.; Gao, Y.; Li, D.; Ma, S. Exploring Fluorite Veins Using Multi-source Remote Sensing Satellite Data: A Case Study from the Shuitou Fluorite Deposit in Inner Mongolia, China. *Northwest. Geol.* **2024**, *57*, 121–134. (In Chinese with English Abstract)
- Shen, J.; Pei, Q.; Zhang, L.; Han, S.; Ma, S.; Chen, L.; Wang, S. Multivariate geological information extraction and prospecting prediction in the Daxigou orefield, Harqin Banner, utilizing SPOT-5, Landsat-9, and ASTER imagery. *Geol. Rev.* **2025**, *71*, 1774–1786. (In Chinese with English Abstract)
- Tsangaratos, P.; Vakalas, I.; Zandarini, I. Distinguishing Lithofacies of Flysch Formations Using Deep Learning Models: Integrating Remote Sensing Data with Morphological Indexes. *Remote Sens.* **2025**, *17*, 422. [\[CrossRef\]](#)
- Feng, Y.; Wang, Y.; Cheng, Y.; Zeng, Z.; Zhang, Y. Refined mapping of granitic pegmatite in the Altay region based on high-resolution remote sensing imagery. *Earth Sci. Inform.* **2026**, *19*, 32. [\[CrossRef\]](#)
- Abrams, M.; Yamaguchi, Y. Twenty Years of ASTER Contributions to Lithologic Mapping and Mineral Exploration. *Remote Sens.* **2019**, *11*, 1394. [\[CrossRef\]](#)
- Appiah-Twum, M.; Xu, W.; Sunkari, E.D. Mapping Lithology with Hybrid Attention Mechanism–Long Short-Term Memory: A Hybrid Neural Network Approach Using Remote Sensing and Geophysical Data. *Remote Sens.* **2024**, *16*, 4613. [\[CrossRef\]](#)
- Yin, C.; Long, Y.; Liu, L.; Khalil, Y.S.; Ye, S. Mapping Ni-Cu-Platinum Group Element-Hosting, Small-Sized, Mafic-Ultramafic Rocks Using WorldView-3 Images and a Spatial-Spectral Transformer Deep Learning Method. *Econ. Geol.* **2024**, *119*, 665–680. [\[CrossRef\]](#)
- Zhang, T.; Zhao, Z.; Dong, P.; Tang, B.-H.; Zhang, G.; Feng, L.; Zhang, X. Rapid lithological mapping using multi-source remote sensing data fusion and automatic sample generation strategy. *Int. J. Digit. Earth* **2024**, *17*, 2420824. [\[CrossRef\]](#)
- Zhang, T.; Zhao, Z.; Zeng, M.; Yang, H. Fast production of large lithologic maps using lithologic sample generation strategies based on 3D spatial positioning. *Geosci. Front.* **2026**, *17*, 102217. [\[CrossRef\]](#)
- Cracknell, M.J.; Reading, A.M. Geological mapping using remote sensing data: A comparison of five machine learning algorithms, their response to variations in the spatial distribution of training data and the use of explicit spatial information. *Comput. Geosci.* **2014**, *63*, 22–33. [\[CrossRef\]](#)
- Bahrami, H.; Esmaeili, P.; Homayouni, S.; Pour, A.B.; Chokmani, K.; Bahroudi, A. Machine Learning-Based Lithological Mapping from ASTER Remote-Sensing Imagery. *Minerals* **2024**, *14*, 202. [\[CrossRef\]](#)
- Ghoneim, S.M.; Hamimi, Z.; Abdelrahman, K.; Khalifa, M.A.; Shabban, M.; Abdelmaksoud, A.S. Machine learning and remote sensing-based lithological mapping of the Duwi Shear-Belt area, Central Eastern Desert, Egypt. *Sci. Rep.* **2024**, *14*, 17010. [\[CrossRef\]](#) [\[PubMed\]](#)
- Hajaj, S.; El Harti, A.; Pour, A.B.; Jellouli, A.; Adiri, Z.; Hashim, M. A review on hyperspectral imagery application for lithological mapping and mineral prospecting: Machine learning techniques and future prospects. *Remote Sens. Appl. Soc. Environ.* **2024**, *35*, 101218. [\[CrossRef\]](#)
- Liu, L.; Yin, C.; Wen, H.; Zhang, G. Mapping lithium hosting claystone using WorldView-3 image and deep learning method. *Gondwana Res.* **2026**, *151*, 18–35. [\[CrossRef\]](#)
- Xie, F.; Yang, Y. Hierarchical Multiscale Fusion with Coordinate Attention for Lithologic Mapping from Remote Sensing. *Remote Sens.* **2026**, *18*, 413. [\[CrossRef\]](#)
- Kumar, C.; Chatterjee, S.; Oommen, T.; Guha, A. Automated lithological mapping by integrating spectral enhancement techniques and machine learning algorithms using AVIRIS-NG hyperspectral data in Gold-bearing granite-greenstone rocks in Hutti, India. *Int. J. Appl. Earth Obs. Geoinf.* **2020**, *86*, 102006. [\[CrossRef\]](#)
- Wang, Z.; Zuo, R.; Dong, Y. Mapping of Himalaya Leucogranites Based on ASTER and Sentinel-2A Datasets Using a Hybrid Method of Metric Learning and Random Forest. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *13*, 1925–1936. [\[CrossRef\]](#)
- Lu, J.; Han, L.; Liu, L.; Wang, J.; Xia, Z.; Jin, D.; Zha, X. Lithology classification in semi-arid area combining multi-source remote sensing images using support vector machine optimized by improved particle swarm algorithm. *Int. J. Appl. Earth Obs. Geoinf.* **2023**, *119*, 103318. [\[CrossRef\]](#)
- Liu, H.; Wu, K.; Xu, H.; Xu, Y. Lithology Classification Using TASI Thermal Infrared Hyperspectral Data with Convolutional Neural Networks. *Remote Sens.* **2021**, *13*, 3117. [\[CrossRef\]](#)
- Wang, X.; Zuo, R.; Wang, Z. Lithological Mapping Using a Convolutional Neural Network based on Stream Sediment Geochemical Survey Data. *Nat. Resour. Res.* **2022**, *31*, 2397–2412. [\[CrossRef\]](#)
- Zheng, S.; Zhou, Y.; An, Y.; Cui, X.; Shi, P. Granitoid Mapping with Convolutional Neural Network from ASTER and Landsat 8 OLI Data: A Case Study in the Western Junggar Orogen. *Remote Sens.* **2025**, *17*, 384. [\[CrossRef\]](#)

24. Wu, S.; Liu, Y. Interpretable Dual-Channel Convolutional Neural Networks for Lithology Identification Based on Multisource Remote Sensing Data. *Remote Sens.* **2025**, *17*, 1314. [\[CrossRef\]](#)
25. Chen, L.; Zhang, N.; Chang, J.; Liao, S.; Tao, J.; Zhang, H.; Li, S. Indicator element selection and lithological mapping using deep learning methods in the Dahongliutan area, NW China. *J. Geochem. Explor.* **2025**, *268*, 107611. [\[CrossRef\]](#)
26. Ye, B.; Tian, S.; Cheng, Q.; Ge, Y. Application of Lithological Mapping Based on Advanced Hyperspectral Imager (AHSI) Imagery Onboard Gaofen-5 (GF-5) Satellite. *Remote Sens.* **2020**, *12*, 3990. [\[CrossRef\]](#)
27. Liu, L.; Yin, C.; Shaheen Khalil, Y.; Hong, J.; Feng, J.; Zhang, H. Alteration Mapping for Porphyry Cu Targeting in the Western Chagai Belt, Pakistan, Using ZY1-02D Spaceborne Hyperspectral Data. *Econ. Geol.* **2024**, *119*, 331–353. [\[CrossRef\]](#)
28. Liu, L.; Feng, J.; Han, L.; Zhou, J.; Xu, X.; Liu, R.; Li, R. Mineral mapping using spaceborne Tiangong-1 hyperspectral imagery and ASTER data: A case study of alteration detection in support of regional geological survey at Jintanzi-Malianquan area, Beishan, Gansu Province, China. *Geol. J.* **2018**, *53*, 372–383. [\[CrossRef\]](#)
29. Dong, Y.; Yang, Z.; Liu, Q.; Zuo, R.; Wang, Z. Fusion of GaoFen-5 and Sentinel-2B data for lithological mapping using vision transformer dynamic graph convolutional network. *Int. J. Appl. Earth Obs. Geoinf.* **2024**, *129*, 103780. [\[CrossRef\]](#)
30. Wang, S.; Han, W.; Huang, X.; Zhang, X.; Wang, L.; Li, J. Trustworthy remote sensing interpretation: Concepts, technologies, and applications. *ISPRS J. Photogramm. Remote Sens.* **2024**, *209*, 150–172. [\[CrossRef\]](#)
31. Chen, X.; Liu, Y.; Han, W.; Zheng, X.; Wang, S.; Wang, J.; Wang, L. A vision-language foundation model-based multi-modal retrieval-augmented generation framework for remote sensing lithological recognition. *ISPRS J. Photogramm. Remote Sens.* **2025**, *225*, 328–340. [\[CrossRef\]](#)
32. Wang, Z.; Tian, S. Ground object information extraction from hyperspectral remote sensing images using deep learning algorithm. *Microprocess. Microsyst.* **2021**, *87*, 104394. [\[CrossRef\]](#)
33. Coleman, R.G. Continental growth of northwest China. *Tectonics* **1989**, *8*, 621–635. [\[CrossRef\]](#)
34. Song, S.; Niu, Y.; Su, L.; Xia, X. Tectonics of the North Qilian orogen, NW China. *Gondwana Res.* **2013**, *23*, 1378–1401. [\[CrossRef\]](#)
35. Wu, C.; Zuza, A.V.; Yin, A.; Liu, C.; Reith, R.C.; Zhang, J.; Liu, W.; Zhou, Z. Geochronology and geochemistry of Neoproterozoic granitoids in the central Qilian Shan of northern Tibet: Reconstructing the amalgamation processes and tectonic history of Asia. *Lithosphere* **2017**, *9*, 609–636. [\[CrossRef\]](#)
36. Xiong, X.; Li, H.; Chen, X.; Li, R.; Zhou, J.; Gao, R.; Eccles, J.D. Subduction retreat recorded in the southern Beishan orogenic collage of the Central Asian Orogenic Belt: New evidence from deep seismic reflection profiling. *Geol. Soc. Am. Bull.* **2024**, *137*, 1254–1266. [\[CrossRef\]](#)
37. Yun, L.; Wang, J.; Zhang, J.; Zhang, H.; Dong, Z. Late Quaternary tectonics of the Jiujiang-Bantan fault along the southern Beishan Block and its implication for the northward growth of the Tibetan Plateau. *J. Struct. Geol.* **2025**, *198*, 105458. [\[CrossRef\]](#)
38. Tapponnier, P.; Zhiqin, X.; Roger, F.; Meyer, B.; Arnaud, N.; Wittlinger, G.; Jingsui, Y. Oblique Stepwise Rise and Growth of the Tibet Plateau. *Science* **2001**, *294*, 1671–1677. [\[CrossRef\]](#) [\[PubMed\]](#)
39. Liu, C.; Ji, L.; Zhu, L.; Xu, C.; Qiu, J. Interseismic strain rate distribution model of the Altyn Tagh Fault constrained by InSAR and GPS. *Earth Planet. Sci. Lett.* **2024**, *642*, 118884. [\[CrossRef\]](#)
40. Pinzon, N.; Klinger, Y.; Xu, X.; Tapponnier, P.; Jing, L.Z.; Van Der Woerd, J.; Li, K.; Gao, M. Spatiotemporal Clustering of Large Earthquakes Along the Central-Eastern Sections of the Altyn Tagh Fault, China. *J. Geophys. Res. Solid Earth* **2024**, *129*, e2024JB028912. [\[CrossRef\]](#)
41. Gan, B.; Diwu, C.; Wang, T.; Pei, Q.; Wang, S. Early Paleozoic intermediate igneous rocks in the Dunhuang Block, southern Central Asian Orogenic Belt: Geochronology, petrogenesis, and implications for tectonic evolution. *J. Geol. Soc.* **2022**, *180*, jgs2021-178. [\[CrossRef\]](#)
42. Wu, C.; Chen, X.; Ding, L. Tectonic evolution and Cenozoic deformation history of the Qilian orogen. *Earth Sci. Front.* **2023**, *33*, 262–281.
43. Gan, B.; Tang, J.; Safonova, I.; Qin, L.; Diwu, C. Devonian continental arc magmatism in the southern Central Asian Orogenic Belt: Evidence from the Dunhuang Block, NW China. *Geosci. Front.* **2023**, *14*, 101643. [\[CrossRef\]](#)
44. Cao, H.-W.; Pei, Q.-M.; Santosh, M.; Li, G.-M.; Zhang, L.-K.; Zhang, X.-F.; Zhang, Y.-H.; Zou, H.; Dai, Z.-W.; Lin, B.; et al. Himalayan leucogranites: A review of geochemical and isotopic characteristics, timing of formation, genesis, and rare metal mineralization. *Earth-Sci. Rev.* **2022**, *234*, 104229. [\[CrossRef\]](#)
45. Bentahar, I.; Raji, M. Comparison of Landsat OLI, ASTER, and Sentinel 2A data in lithological mapping: A Case study of Rich area (Central High Atlas, Morocco). *Adv. Space Res.* **2021**, *67*, 945–963. [\[CrossRef\]](#)
46. Gemusse, U.; Cardoso-Fernandes, J.; Lima, A.; Teodoro, A. Identification of pegmatites zones in Muiane and Naipa (Mozambique) from Sentinel-2 images, using band combinations, band ratios, PCA and supervised classification. *Remote Sens. Appl. Soc. Environ.* **2023**, *32*, 101022. [\[CrossRef\]](#)
47. Arian, H.; Alaminia, Z.; Ahmadi, H.; Pour, A.B.; Tabatabaei Manesh, S.M.; Lentz, D.R.; Parsa, L. Mapping potential zones of Lithium-bearing pegmatites using ASTER and Sentinel-2 remote sensing imagery in the Tagablor Pegmatite Field, Central Afghanistan. *Remote Sens. Appl. Soc. Environ.* **2025**, *38*, 101625. [\[CrossRef\]](#)

48. Liu, L.; Zhou, J.; Yin, F.; Feng, M.; Zhang, B. The reconnaissance of mineral resources through aster data-based image processing, interpreting and ground inspection in the Jiafushaersu area, West Junggar, China. *J. Earth Sci.* **2014**, *25*, 397–406. [\[CrossRef\]](#)
49. Yang, M.; Kang, L.; Chen, H.; Zhou, M.; Zhang, J. Lithological mapping of East Tianshan area using integrated data fused by Chinese GF-1 PAN and ASTER multi-spectral data. *Open Geosci.* **2018**, *10*, 532–543. [\[CrossRef\]](#)
50. Khotimah, W.N.; Bennamoun, M.; Boussaid, F.; Sohel, F.; Edwards, D. A High-Performance Spectral-Spatial Residual Network for Hyperspectral Image Classification with Small Training Data. *Remote Sens.* **2020**, *12*, 3137. [\[CrossRef\]](#)
51. Zhu, M.; Jiao, L.; Liu, F.; Yang, S.; Wang, J. Residual Spectral-Spatial Attention Network for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 449–462. [\[CrossRef\]](#)
52. Liu, B.; Yu, X.; Yu, A.; Zhang, P.; Wan, G.; Wang, R. Deep Few-Shot Learning for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 2290–2304. [\[CrossRef\]](#)
53. Mou, L.; Ghamisi, P.; Zhu, X.X. Unsupervised Spectral–Spatial Feature Learning via Deep Residual Conv–Deconv Network for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 391–406. [\[CrossRef\]](#)
54. Hu, S.; Li, H. Hyperspectral Unmixing with Multi-Scale Convolution Attention Network. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2024**, *17*, 2531–2542. [\[CrossRef\]](#)
55. Mohla, S.; Pande, S.; Banerjee, B.; Chaudhuri, S. FusAtNet: Dual Attention based SpectroSpatial Multimodal Fusion Network for Hyperspectral and LiDAR Classification. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Seattle, WA, USA, 14–19 June 2020; pp. 416–425. [\[CrossRef\]](#)
56. Zhang, H.; Liu, H.; Yang, R.; Wang, W.; Luo, Q.; Tu, C. Hyperspectral Image Classification Based on Double-Branch Multi-Scale Dual-Attention Network. *Remote Sens.* **2024**, *16*, 2051. [\[CrossRef\]](#)
57. Han, W.; Zhang, X.; Wang, Y.; Wang, L.; Huang, X.; Li, J.; Wang, S.; Chen, W.; Li, X.; Feng, R.; et al. A survey of machine learning and deep learning in remote sensing of geological environment: Challenges, advances, and opportunities. *ISPRS J. Photogramm. Remote Sens.* **2023**, *202*, 87–113. [\[CrossRef\]](#)
58. Sabale, S.P.; Jadhav, C.R. Hyperspectral Image Classification Methods in Remote Sensing-A Review. In Proceedings of the 1st International Conference on Computing Communication Control and Automation ICCUBEA, Pune, India, 26–27 February 2015; pp. 679–683. [\[CrossRef\]](#)
59. Xu, X.; Li, W.; Ran, Q.; Du, Q.; Gao, L.; Zhang, B. Multisource Remote Sensing Data Classification Based on Convolutional Neural Network. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 937–949. [\[CrossRef\]](#)
60. Jung, S.; Dagobert, T.; Morel, J.-M.; Facciolo, G. A Review of t-SNE. *Image Process. Line* **2024**, *14*, 250–270. [\[CrossRef\]](#)
61. Melit Devassy, B.; George, S.; Nussbaum, P. Unsupervised Clustering of Hyperspectral Paper Data Using t-SNE. *J. Imaging* **2020**, *6*, 29. [\[CrossRef\]](#) [\[PubMed\]](#)
62. Wang, Z.; Zuo, R. An Evaluation of Convolutional Neural Networks for Lithological Mapping Based on Hyperspectral Images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2024**, *17*, 6414–6425. [\[CrossRef\]](#)
63. Indra, D.; Fadlillah, H.M.; Kasman; Ilmawan, L.B. Rice Texture Analysis Using GLCM Features. In Proceedings of the 2021 International Conference on Electrical, Computer and Energy Technologies (ICECET), Cape Town, South Africa, 9–10 December 2021; pp. 1–5. [\[CrossRef\]](#)
64. Lu, Y.; Yang, C. Influence of GLCM texture parameters on lithological mapping using Sentinel-1 imagery. *Geocarto Int.* **2024**, *39*, 2425183. [\[CrossRef\]](#)
65. Fasihi, M.; Falcon, A.; Alberti, G.; Cadez, L.; Giannetti, F.; Tomao, A.; Serra, G. Evaluating organic carbon in living and dead trees using GLCM features and explainable machine learning: Insights from Italian national forest. *Ann. GIS* **2025**, *31*, 651–677. [\[CrossRef\]](#)
66. Pal, M.; Rasmussen, T.; Porwal, A. Optimized Lithological Mapping from Multispectral and Hyperspectral Remote Sensing Images Using Fused Multi-Classifiers. *Remote Sens.* **2020**, *12*, 177. [\[CrossRef\]](#)
67. Li, P.; Yu, H.; Cheng, T. Lithologic mapping using ASTER imagery and multivariate texture. *Can. J. Remote Sens.* **2009**, *35*, S117–S125. [\[CrossRef\]](#)

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.