

Article

Method of Convolutional Neural Networks for Lithological Classification Using Multisource Remote Sensing Data

Zixuan Zhang, Yuanjin Xu *  and Jianguo Chen 

School of Earth Resources, China University of Geosciences, Wuhan 430074, China; zzx85@cug.edu.cn (Z.Z.); jgchen@cug.edu.cn (J.C.)

* Correspondence: xuyj@cug.edu.cn; Tel.: +86-27-67883066

Highlights

What are the main findings?

- A dual-branch deep learning framework (TSFDNet) was proposed to integrate Sentinel-1 microwave and GF-5 hyperspectral data, enabling accurate lithological classification in densely vegetated areas.
- Three functional modules were designed—Texture Feature Extraction Module (TFEM), Spatial-Spectral Feature Module (SSFM), and Feature Fusion Extraction Module (FFEM)—to efficiently extract and fuse multisource features, achieving an overall accuracy of 97.24% and a mean IoU of 90.43%.

What are the implications of the main findings?

- This study provides a deep learning framework based on adaptive fusion of multi-source remote sensing data, effectively enhancing lithological discrimination under vegetation-covered and complex surface conditions.

Abstract

Xinfeng County, Shaoguan City, Guangdong Province, China, is a typical vegetation-covered area that suffers from severe attenuation of rock and mineral spectral information in remote sensing images owing to dense vegetation. This situation limits the accuracy of traditional lithological mapping methods, making them unable to meet geological mapping demands under complex conditions, and thus necessitating a tailored lithological identification model. To address this issue, in this study, the penetration capability of microwave remote sensing (for extracting indirect textural features of lithology) was combined with the spectral superiority of hyperspectral remote sensing (for capturing lithological spectral features), resulting in a dual-branch deep-learning framework for lithological classification based on multisource remote sensing data. The framework independently extracts features from Sentinel-1 imagery and Gaofen-5 data, integrating three key modules: texture feature extraction, spatial-spectral feature extraction, and attention-based adaptive feature fusion, to realize deep and efficient fusion of heterogeneous remote sensing information. Ablation and comparative experiments were conducted to evaluate each module's contribution. The results show that the dual-branch architecture effectively captures the complementary and discriminative characteristics of multimodal data, and that the encoder-decoder structure demonstrates strong robustness under complex conditions such as dense vegetation. The final model achieved 97.24% overall accuracy and 90.43% mean intersection-over-union score, verifying its effectiveness and generalizability in complex geological environments. The proposed multi-source remote sensing-based lithological classification model overcomes the limitations of single-source data by integrating indirect lithological texture features



Academic Editor: Lenio Soares Galvao

Received: 29 October 2025

Revised: 14 December 2025

Accepted: 19 December 2025

Published: 22 December 2025

Copyright: © 2025 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.

containing vegetation structural information with spectral features, thereby providing a viable approach for lithological mapping in vegetated regions.

Keywords: deep learning; dual-branch network; lithological classification; microwave data; vegetation cover

1. Introduction

Lithological mapping, a fundamental component of geological surveys, plays a crucial role in revealing the structural features and stratigraphic distributions of geological formations [1–6]. With rapid advancements in remote-sensing technologies, particularly hyperspectral and microwave remote sensing, new technical pathways have emerged for lithological mapping.

Hyperspectral remote sensing uses hundreds of contiguous spectral bands, endowing it with strong material identification capabilities. It demonstrates significant advantages for extracting the spectral features of minerals and lithologies [7,8]. This technology has been widely applied in geological mapping and mineral resource exploration and has yielded numerous effective results [9–16]. Compared to traditional multispectral remote sensing, hyperspectral data perform better in fine-scale surface classification. They are particularly superior in identifying rocks with similar visual appearances but distinct spectral characteristics [17–19]. However, the relatively low spatial resolution and high dimensionality of spectral information often hinder its full utilization in practical applications, leaving room for further improvements in classification accuracy.

Microwave remote sensing features a strong penetration ability and all-weather imaging capabilities. It offers rich textural information on lithological surfaces and serves as an important complement to hyperspectral data [20,21]. Microwave data can capture both surface and shallow subsurface structural features of rocks, and can be used to distinguish lithological types by analyzing backscattering characteristics. Researchers have leveraged backscatter coefficients, polarimetric decomposition parameters, and textural features derived from gray-level co-occurrence matrices (GLCMs) for lithological classification [22–25]. Furthermore, microwave remote sensing can provide high-resolution digital elevation models, which are particularly valuable for lithological discrimination in complex terrain [26]. However, microwave data are often subject to high acquisition costs and complex preprocessing procedures. More importantly, when used alone for lithology identification, these methods can only provide limited spatial texture information and cannot provide mineral spectral information, resulting in unsatisfactory recognition accuracy. [27].

Single-source data often struggle to achieve accurate classification under complex surface conditions, particularly in areas with vegetation cover. Considering the complementary strengths and inherent limitations of hyperspectral and microwave remote sensing, multisource data fusion has emerged as a logical and effective solution. This approach integrates information from multiple sensors, with each contributing distinct yet synergistic features to comprehensively characterize the same scene [23]. Fusion approaches enhance the accuracy and robustness of lithological classifications across spectral, spatial, and textural dimensions by jointly exploiting the strengths of hyperspectral and microwave data. Recently, multisource fusion has shown great promise for image classification, geological mapping, and structural analysis [28].

Fusion techniques are generally categorized into four types: pixel-level, feature-level, decision-level, and deep-learning (DL)-based multimodal fusion techniques [29,30]. Pixel-level methods such as component substitution and high-pass filtering directly construct

new fused images. Feature-level fusion extracts textural, spectral, or statistical features from different sensors to form composite feature datasets for downstream tasks and is widely used in lithological classification [31,32]. Decision-level fusion combines the outputs from multiple classifiers using rule-based or ensemble strategies, such as decision trees or classifier fusion [33]. With the proliferation of DL technologies, neural-network-based fusion frameworks have garnered significant attention. Attention mechanisms are increasingly employed in these frameworks to enhance intermodal feature interactions and improve fusion effectiveness [34].

Despite these advances, current multisource fusion methods have notable limitations in terms of lithological classification. Several approaches rely on simple feature concatenation and fail to exploit deeper semantic information from heterogeneous modalities [35]. However, even when DL architectures are adopted, some models overlook semantic discrepancies between modalities during the fusion process. This shortcoming leads to insufficient representation of lithological classes under complex conditions, particularly for distinguishing spectrally similar rock types. Therefore, an urgent need exists for multisource fusion frameworks with strong modality-specific feature extraction and semantic alignment capabilities. Such frameworks can fully harness the complementarity of hyperspectral and microwave data, and enhance lithological recognition in challenging surface environments.

To address these challenges, this paper proposes a DL-based dual-branch multisource remote sensing network for lithological classification. The model adopts a parallel dual-branch encoder–decoder architecture to independently extract deep features from hyperspectral and microwave data. Three key modules are introduced to accommodate the distinct representation characteristics of each modality: a texture feature extraction module (TFEM) for mining deep structural features from microwave data, a spatial–spectral feature extraction module (SSFM) for capturing hyperspectral information, and an adaptive feature fusion module (FFEM) that assigns weights to different features on the basis of learned attention mechanisms to achieve optimal fusion. With Xinfeng County in Shaoguan City, Guangdong Province, China as the study area, both ablation and comparative experiments were conducted to systematically evaluate the contributions of each module. The results demonstrate that the proposed model significantly outperforms traditional DL methods in terms of lithological classification and achieves the highest classification accuracy.

2. Network for Lithological Classification

2.1. General Framework of the Network

The proposed dual-branch multisource feature fusion network, called the dual-branch texture and spectral feature network (TSDNet), is based on the ERFNet [36] architecture. The network effectively extracts deep features from each data modality, including indirect lithological texture features and spectral features, and assigns appropriate importance to different types of features. By integrating both texture and spectral information in a complementary manner, it significantly enhances the accuracy of lithological classification. The overall structure of the system is shown in Figure 1.

Building on ERFNet’s original encoder–decoder framework, TSFDNet incorporates three additional modules: a texture feature extraction module (TFEM), spatial–spectral feature module (SSFM), and feature fusion extraction module (FFEM). The TFEM extracts the microwave data features using an enhanced attention mechanism. The SSFM employs a triple-branch structure to capture both spectral and spatial characteristics from hyperspectral imagery. The FFEM adopts a gated unit to adaptively fuse microwave and hyperspectral features by learning their weight distributions, enabling optimal integration of the two modalities.

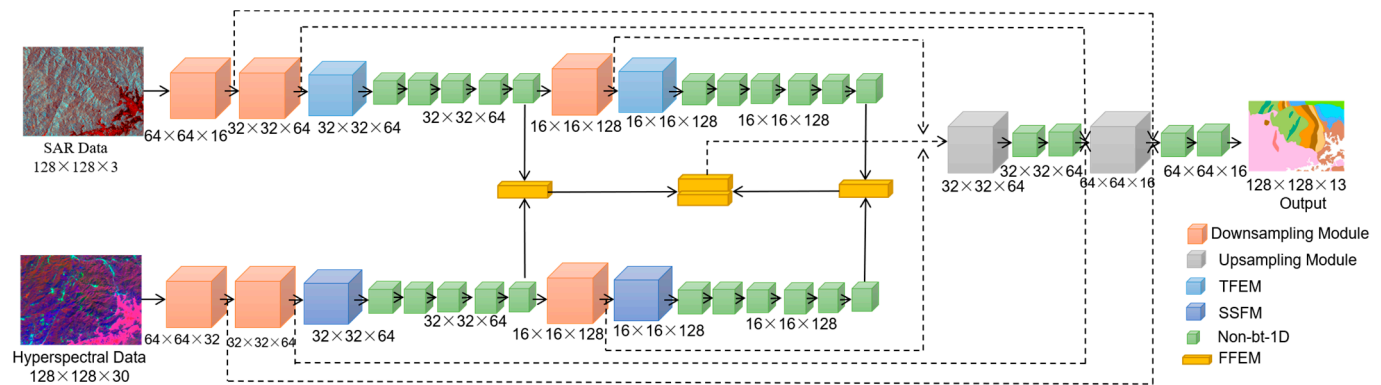


Figure 1. Dual-branch texture and spectral feature network. Among them, the red module is for downsampling, the gray module is for upsampling, the light blue module is for texture feature extraction, the dark blue module is for spectral feature extraction, the orange module is for feature fusion, and the green module is the basic module of ERFNet.

ERFNet is an efficient convolutional neural network designed for real-time semantic segmentation. Semantic segmentation has achieved notable results in image recognition, particularly for identifying lithologies, geological structures, and related features [34,37,38]. Its non-bottleneck-1D (Non-bt-1D) module uses cascaded 3×1 and 1×3 convolutional kernels for successive operations. This design significantly reduces computational complexity while preserving the spatial resolution of the feature map. The original ERFNet modular structure and parameter configurations, including the downsampling and upsampling layers, were fully retained in the proposed network to ensure performance stability and reproducibility.

2.2. Texture Feature Extraction Module

Microwave image texture is typically characterized by GLCM parameters, such as entropy, contrast, correlation, and homogeneity [29,39]. However, in densely vegetated areas, such textural features become obscure and difficult to represent using conventional GLCM-based descriptors. Self-attention mechanisms have demonstrated notable advantages in capturing complex surface patterns from microwave data [40]. Consequently, we propose a TFEM based on an enhanced self-attention mechanism to extract and represent indirect texture of lithology efficiently (Figure 2).

This module adopts a multilevel feature fusion strategy. First, channel shuffle operations reorganize the feature channels, followed by the generation of query (Q), key (K), and value (V) matrices. The query feature Q is processed through a dual-attention mechanism, where spatial attention (SA) and channel attention (CA) capture spatial and channel-wise contextual information, respectively. The key feature K undergoes hierarchical convolution comprising a $k \times k$ convolutional layer, batch normalization, and rectified linear unit (ReLU) activation to enhance the local texture patterns from the microwave data. This design improves local texture extraction and reduces computational complexity.

Finally, the global contextual and localized texture features are adaptively fused using a weighted strategy. The fused features are normalized by a sigmoid activation function and then multiplied element-wise by the value feature V to generate the final microwave texture feature representation.

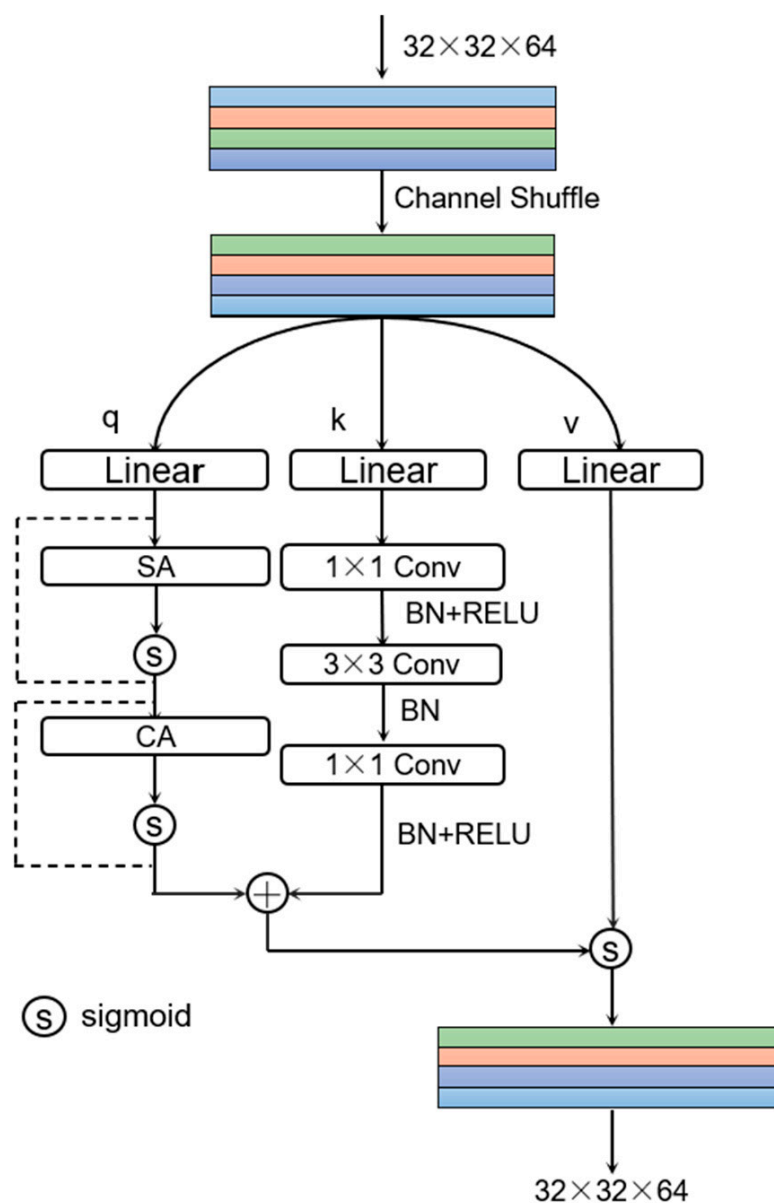


Figure 2. Texture feature extraction module. The module has the same input and output size, both $32 \times 32 \times 64$, where SA represents spatial attention and CA represents channel attention.

2.3. Spatial–Spectral Feature Module

Hyperspectral data contain rich spectral information; however, the local spatial context is often lost during spectral feature extraction [41,42]. To address this issue, the SSFM is proposed, based on multi-feature fusion to efficiently integrate spectral features with spatial contextual information (Figure 3). This module innovatively introduces a local spatial enhancement mechanism into conventional spectral feature-extraction techniques.

The SSFM adopts a three-branch parallel architecture to capture deep spectral features, local spatial features, and long-range dependencies. A residual connection mechanism is employed throughout the module to ensure stability and robustness.

1. The first branch utilizes an enhanced self-attention mechanism to model spectral dependencies. It generates Q, K, and V vectors using 3×3 convolutions with approximate linear transformations and normalization. These features are concatenated and passed through a dilated convolution layer and standard convolution layer to extract deeper spectral features.

2. The second and third branches employ a symmetric dual-branch architecture based on dilated convolutions to enrich the spatial representations and enhance the local perceptual capacity. In these branches, the hyperspectral data are processed using three standard convolutional layers and one dilated convolutional layer. Asymmetric convolution kernels are incorporated into the standard layers to reduce the parameter overhead while maintaining the spatial extraction capability.

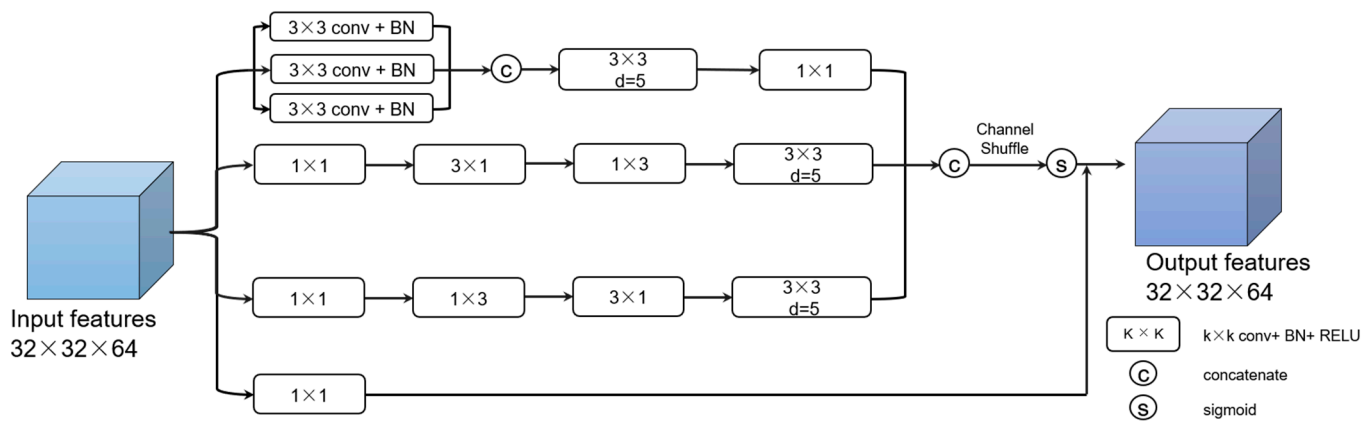


Figure 3. Spatial-spectral feature module. The module has the same input and output size, both $32 \times 32 \times 64$.

Finally, the spectral features from the first branch are fused with the spatial features from the second and third branches. A channel shuffle operation enhances interchannel information interaction. The final spatial–spectral features are obtained through a weighted fusion strategy, which effectively captures both the spectral and spatial characteristics of the hyperspectral data.

2.4. Feature Fusion Extraction Module

To effectively exploit the complementary information learned by the two branches, it is necessary to integrate the features extracted from each branch into a unified representation. Common feature fusion strategies include linear, attention, and gated unit-based methods [43]. Attention-based methods assign varying weights to different regions, allowing the model to focus on informative areas while suppressing redundant or irrelevant regions. This approach has been widely applied to various tasks owing to its effectiveness in enhancing discriminative feature representations. Given the superior capability of efficient multiscale attention (EMA) [44] to extract salient features from individual modalities, an EMA-based FFEM is proposed for feature fusion (Figure 4).

In the FFEM, microwave-derived texture features and hyperspectral-derived spatial–spectral features are first fed into an EMA block to obtain joint aggregated features, and global average pooling is applied separately to the texture features, spatial–spectral features, and their aggregated representation to extract global contextual information and mitigate the influence of spatial dimensionality.

The globally pooled features are normalized using a sigmoid activation function to generate adaptive weights that quantified the feature importance. Subsequently, the Softmax function normalizes the weights across the modalities, ensuring an optimal distribution between the texture and spatial–spectral features. Finally, the weighted texture, spatial–spectral, and aggregated features are combined by element-wise summation to produce the final fused feature representation. This design enables the adaptive integration of heterogeneous information and enhances lithological classification capability.

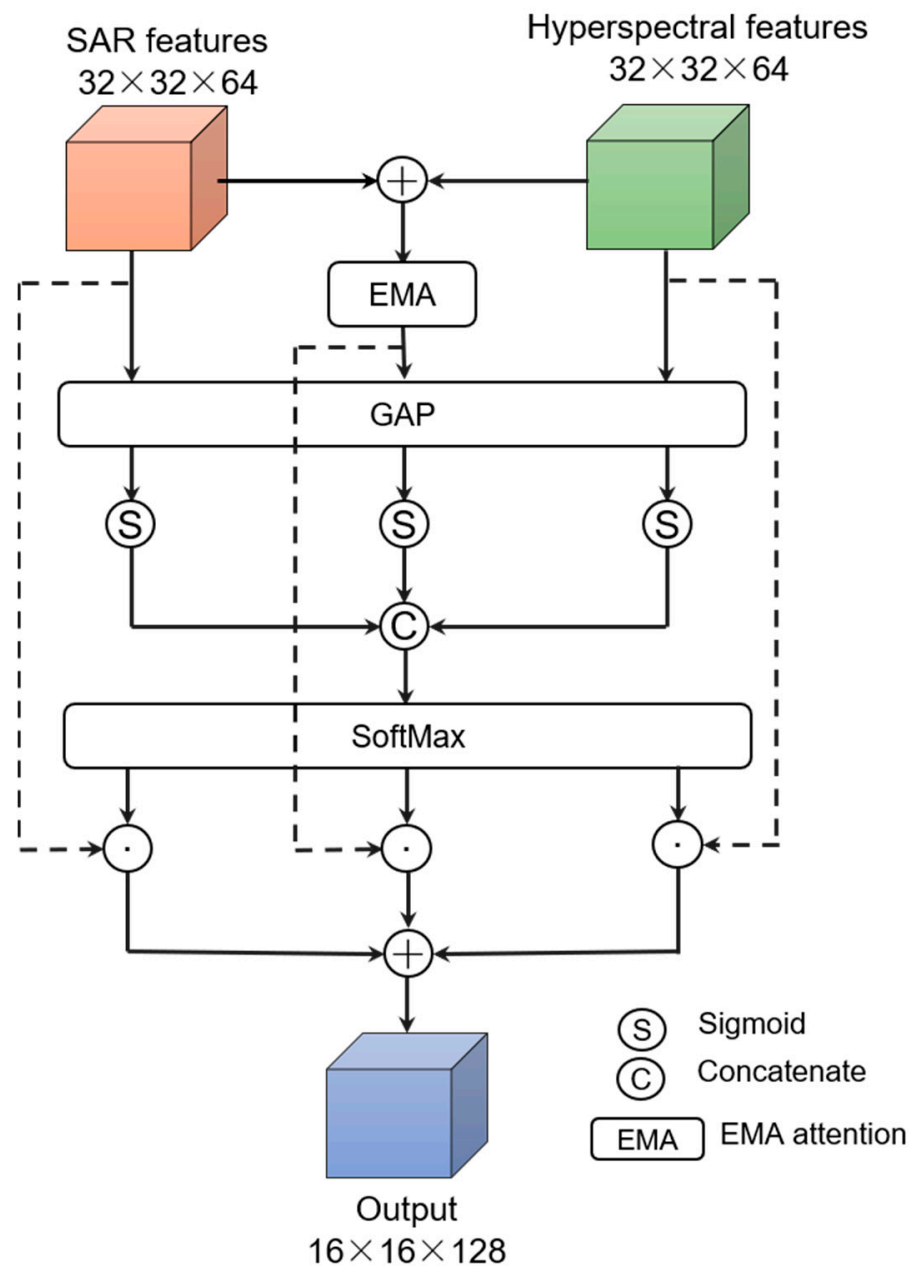


Figure 4. Feature fusion extraction module. Among them, the microwave features and hyperspectral features have the same input size, both $32 \times 32 \times 64$, and the output feature size is $16 \times 16 \times 128$. Solid lines indicate the sequential steps of feature processing. Dashed lines indicate that the features are directly input into the operation pointed to by the arrow without any other processing.

2.5. Indicators for Model Evaluation

To assess TSFDNet performance, several commonly used evaluation metrics were adopted: precision (p), recall (r), F1 score, accuracy, mean intersection over union ($MIoU$), and mean pixel accuracy (MPA) [45,46].

Precision (p) refers to the proportion of correctly predicted lithological categories among those predicted by the model:

$$p = \frac{TP}{TP + FP} \quad (1)$$

Recall (r) indicates the proportion of actual lithological categories in the image that are correctly identified by the model:

$$r = \frac{TP}{TP + FN} \quad (2)$$

The F_1 score represents the harmonic mean of precision and recall, and reaches its maximum when both precision and recall are optimized simultaneously:

$$F_1 = 2 \times \frac{p \times r}{p + r} \quad (3)$$

Accuracy measures the proportion of correctly predicted samples over the total number of samples and is defined as

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

where true positive (TP) is the number of correctly predicted positive samples, true negative (TN) is the number of correctly predicted negative samples, false positive (FP) is the number of incorrectly predicted positive samples, and false negative (FN) is the number of positive samples incorrectly predicted as negative.

$MIoU$ measures the overlap between the predicted segmentation and the ground truth, averaged over all classes. It reflects the consistency between the predictions and reality.

$$MIoU = \frac{1}{N} \sum_{i=1}^N IoU_i \quad (5)$$

MPA is calculated as the average of the per-class pixel accuracies, where the pixel accuracy for class i is defined as the ratio of correctly predicted pixels to the total number of pixels in that class.

$$MPA = \frac{1}{N} \sum_{i=1}^N \frac{TP_i}{TP_i + FN_i} \quad (6)$$

where TP_i and FN_i are the true positives and false negatives, respectively, for class i , and N is the total number of classes.

Higher values of all these metrics indicate better model performance in terms of classification and segmentation accuracy. A detailed introduction to the data in the research area will be presented in the third section.

3. Study Area and Dataset

3.1. Study Area

The study area is located in Xinfeng County, Shaoguan City, Guangdong Province, China, with geographic coordinates ranging from 23°50'00"N to 24°00'00"N and 114°15'00"E to 114°30'00"E. The study area is characterized by a subtropical monsoon climate with high precipitation and humidity, diverse vegetation types, high vegetation coverage, and thick weathering crusts. It has abundant lithological types, including two granite varieties and ten sedimentary rock types, exhibiting the characteristics of a granite-sedimentary rock transition zone. Moreover, it is adjacent to a tin metallogenic belt, rendering the area of great research significance for lithological classification in vegetation-covered regions. The lithology of the study area is mainly granite and different types of sedimentary rocks (Figure 5). The southwestern part of the study area is composed of fine-grained porphyritic granite and coarse-, medium-, and fine-grained monzonitic granites. In the eastern part, ten types of sedimentary rocks are present: siltstone, shale, volcanic rock, muddy silt-

stone, sandstone, coal-bearing clastic rock, micritic limestone, sandstone–siltstone, complex conglomerate, and mudstone.

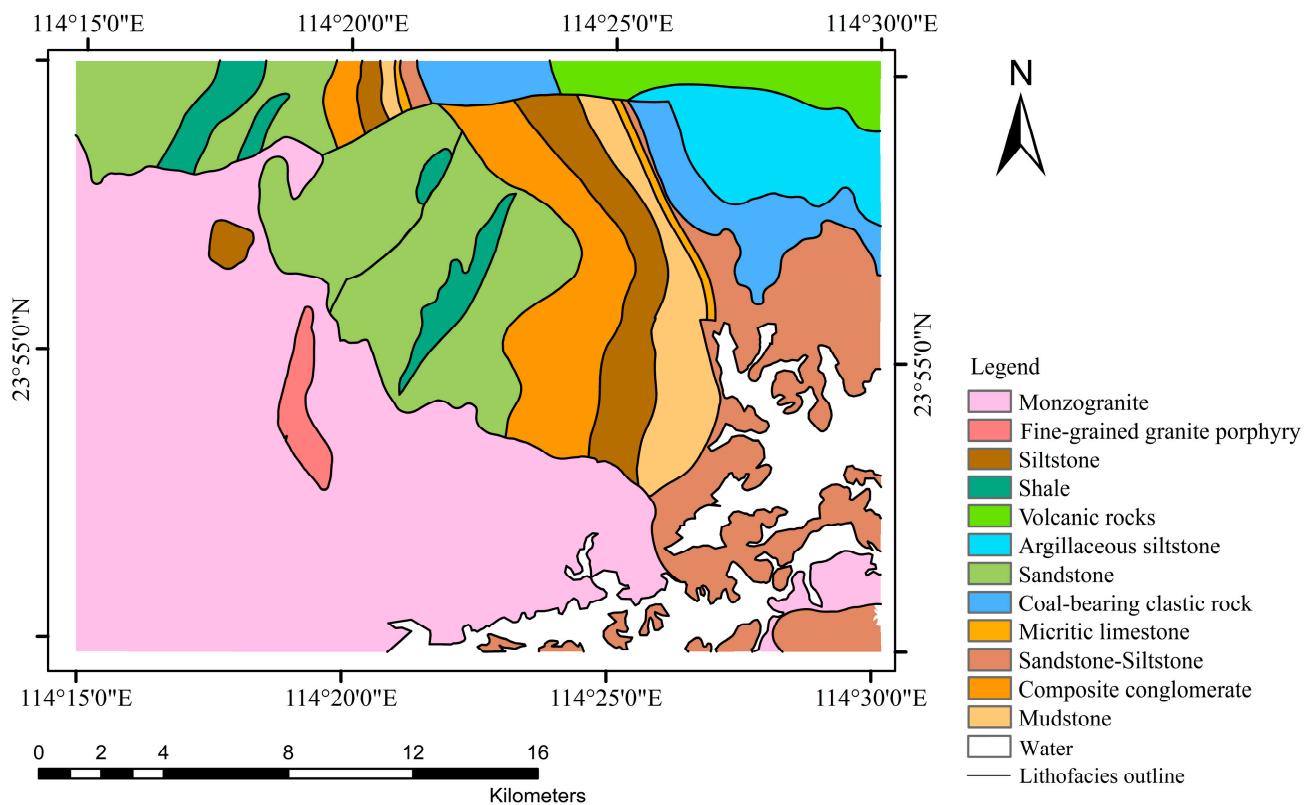


Figure 5. Simplified geological map of the study area.

3.2. Sentinel-1 Imagery and Preprocessing

The Sentinel-1 satellite system, comprising Sentinel-1A and Sentinel-1B, is equipped with a C-band synthetic aperture radar, which enables the acquisition of high-quality surface images under all-weather and all-day conditions.

This feature is needed because variations in soil thickness and mineral weathering products associated with different lithologies result in subtle differences in canopy roughness, slight changes in scattering configurations, and variations in local slopes and surface roughness among the same vegetation types. Although the C-band data of Sentinel-1 cannot penetrate the ground surface in vegetation-covered areas, structural information indirectly related to surface lithology can still be retrieved.

Dual-polarization microwave data released in October 2013 were used in this study. The data belonged to Level-1 ground-range detection products, with a multi-look configuration of 5×1 and a spatial resolution of $10 \text{ m} \times 10 \text{ m}$. Preprocessing steps were performed using the Sentinel application platform software (SNAP 10.0), including orbit correction, radiometric calibration, multi-looking, and geocoding. The processed outputs included the backscattering coefficients (VV/VH/VV-VH). The results are presented in Figure 6.

3.3. GF-5 Imagery and Preprocessing

GF-5, launched on 9 May 2018, is the China's first satellite capable of simultaneously performing integrated observations of both the land and atmosphere. In this study, the GF-5 hyperspectral data released in November 2023 were used. These data were acquired using a visible and shortwave infrared hyperspectral imager with a spatial resolution of $30 \text{ m} \times 30 \text{ m}$, covering 330 spectral bands and a swath width of 60 km.

Preprocessing of the GF-5 data was performed using ENVI 5.6 software, including bad band removal, radiometric calibration, atmospheric correction, and topographic correction. After pre-processing, 228 valid spectral bands were retained. Owing to the significant redundancy among the spectral bands, principal component analysis was performed, and 30 principal components were preserved. The processed GF-5 data consisted of 30 spectral bands (Figure 7). To ensure spatial consistency, the 30-m spatial resolution GF-5 data were resampled to a resolution of 10 m and aligned with the resolution of the Sentinel-1 data. Both datasets were coregistered using the WGS 84 coordinate reference system.

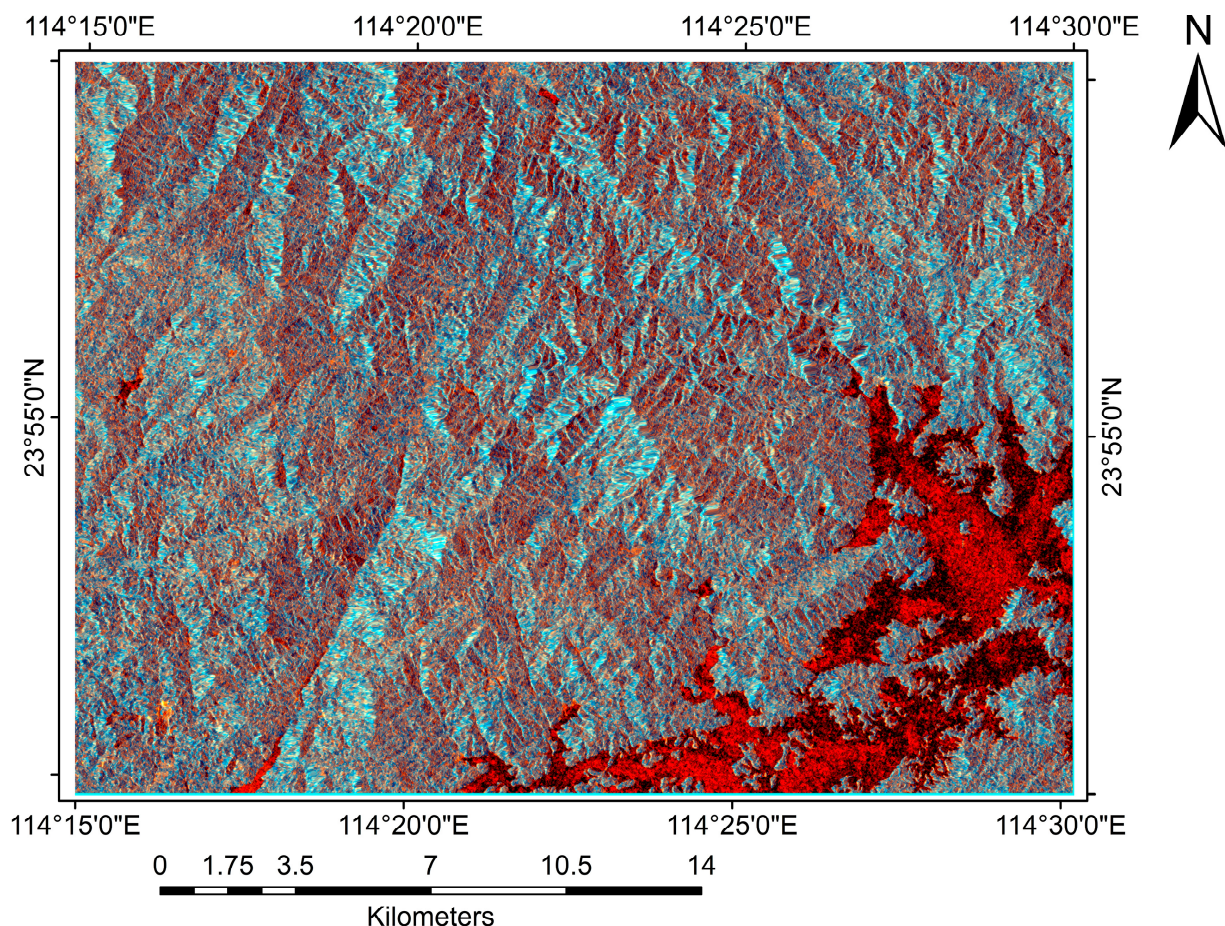


Figure 6. False-color composite of Sentinel-1 data (R: VV, G: VH, B: VV-VH).

3.4. Construction of Geological Labels

The lithological classification of remote sensing images of vegetation-covered areas cannot be reliably achieved by visual interpretation alone. Therefore, the sample labels used in this study were based on geological datasets and field surveys. A 1:50,000-scale geological map was adopted as the primary reference for sample labeling, and the following steps were implemented to ensure label accuracy: First, a 1:200,000-scale geological map was used as a base map to merge the fine-grained geological units in the 1:50,000-scale map, thereby defining a final set of 13 lithological categories. Second, high-resolution Google Earth imagery (0.5–2 m), 30-m topographic elevation data, and expert knowledge were referenced to manually revise erroneous and discontinuous geological boundaries. Finally, field investigations were conducted to verify ambiguous boundaries and delineate the final geological vector lines. Subsequently, the geological vector lines, refined using topological rules, were converted into geological vector polygons, which were then resampled to

a 10-m resolution to generate a raster geological label map with the WGS 84 projection coordinate system.

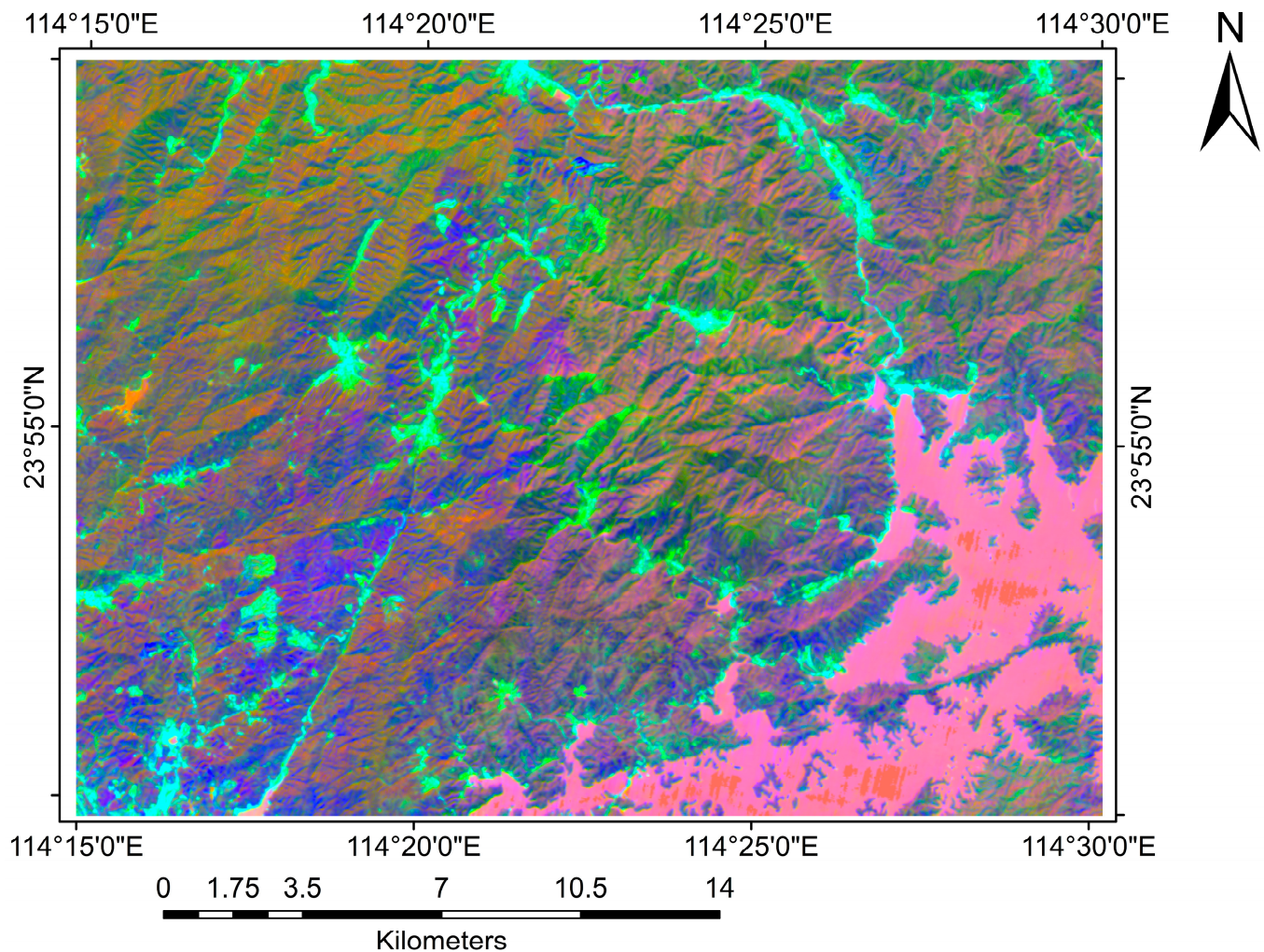


Figure 7. False-color image of principal components of GF-5 (R: first principal component, G: second principal component, B: third principal component).

3.5. Sample Dataset

In this study, the aforementioned remote-sensing data were used to construct a dataset. The label, microwave, and hyperspectral data were cropped into patches of 128×128 pixels, resulting in 1160 images that met the requirements of pixel-level image segmentation tasks. Among these, 80% were randomly selected as the training set and the remaining 20% were used as the validation set.

4. Experimental Results

4.1. Model Training

All the experiments were conducted using an NVIDIA RTX 3070 GPU. The batch size, learning rate, and number of training epochs are critical parameters for ensuring model convergence. The number of training epochs refers to the total number of times the entire dataset is fed through the model during training. The batch size denotes the number of training samples processed simultaneously in each iteration, and the learning rate determines the step size at each iteration while moving toward the minimum of the loss function.

In this experiment, the batch size, learning rate, and number of training epochs were set to 16, 0.0005, and 600, respectively. The model was optimized using a cross-entropy loss function.

As illustrated in Figure 8, both the training and validation loss curves tended to stabilize after approximately 150 epochs. The validation loss converged slightly faster than the training loss, although it exhibited a slightly higher value, which could be attributed to the imbalanced distribution of samples in the validation set. The MIoU of the validation set stabilized at approximately 90% after 150 epochs.

4.2. Ablation Experiments

We conducted ablation experiments based on the original experimental setup to evaluate the effectiveness of the proposed texture feature extraction, spatial-spectral features, and feature fusion modules. Five comparative models were designed in this study.

1. Baseline dual-branch ERFNet: the original dual-branch ERFNet without any of the proposed modules;
2. Texture + spatial-spectral modules: dual-branch ERFNet with both TFEM and SSFM;
3. Texture + fusion modules: dual-branch ERFNet with TFEM and FFEM;
4. Spatial-spectral + fusion modules: dual-branch ERFNet with SSFM and FFEM;
5. TSFDNet (proposed method): Dual-branch ERFNet with all three modules: texture, spatial-spectral, and fusion.

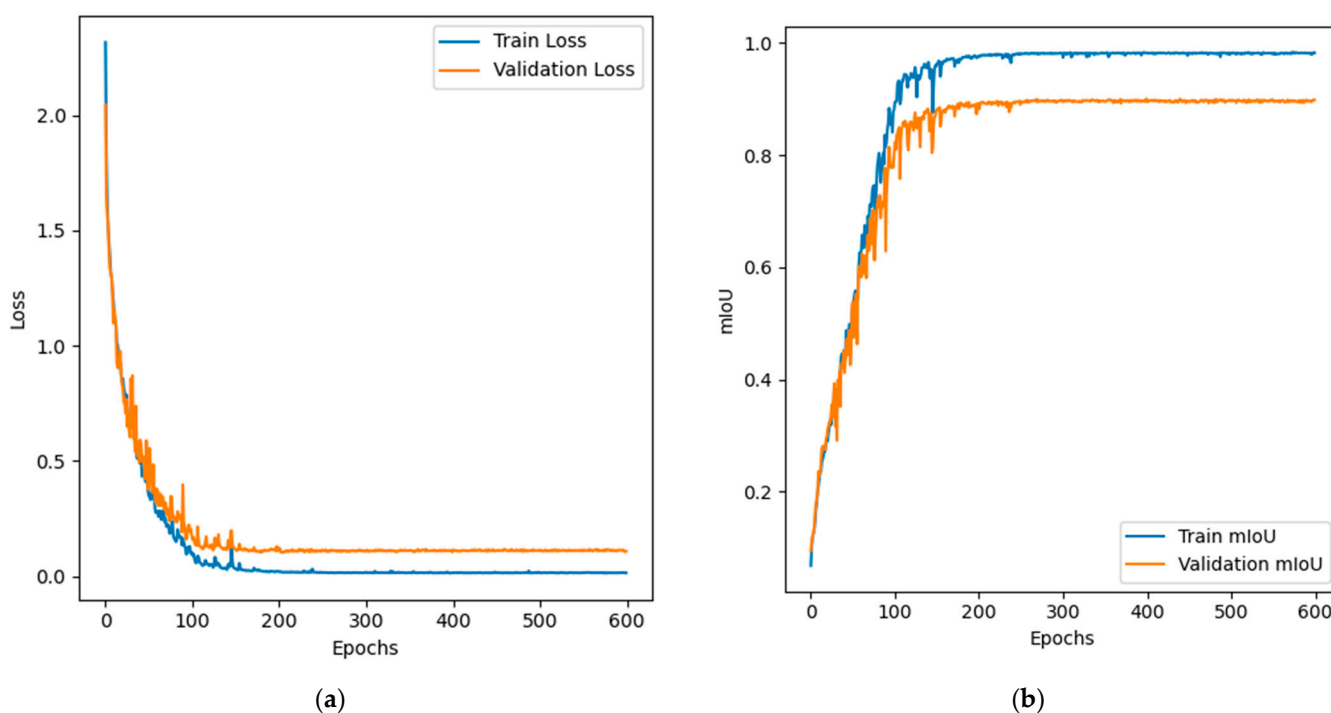


Figure 8. Convergence curves:(a) Loss curve, (b) MIoU curve.

The evaluation metrics for all the models are summarized in Table 1. The full model, which integrates the texture, spatial-spectral, and fusion modules, achieved the highest performance, with an overall accuracy (OA) of 97.24%, MIoU of 90.43%, and MPA of 94.62%. When two modules were combined, the classification performance was lower than that of the full model. The OA for the combinations followed the trend of spatial-spectral + fusion > texture + fusion > texture + spatial-spectral. Finally, when none of the proposed modules were included, the baseline dual-branch ERFNet exhibited the lowest performance, with

an OA reduction of 1.63% compared to the full model, confirming the effectiveness and necessity of the proposed modules in enhancing lithological classification accuracy.

Table 1. Ablation experiment evaluation index results.

Model	Module Combination Options (✓ Indicates Selected)				Evaluation Metrics		
	Dual-Branch ERFNet	TFEM	SSFEM	FFEM	Accuracy	MIoU	MPA
1	✓				95.61%	85.14%	92.73%
2	✓	✓	✓		96.29%	85.37%	92.20%
3	✓	✓		✓	96.35%	85.94%	92.91%
4	✓		✓	✓	97.19%	90.28%	94.60%
5	✓	✓	✓	✓	97.24%	90.43%	94.62%

4.3. Lithological Mapping

The trained model was applied to GF-5 and Sentinel-1 data to predict the distribution of 12 lithological classes within the study area. The classification map generated by the proposed model is presented in Figure 9. Visually, all 12 lithological categories were clearly distinguished and their spatial distribution patterns were closely aligned with known geological structures.

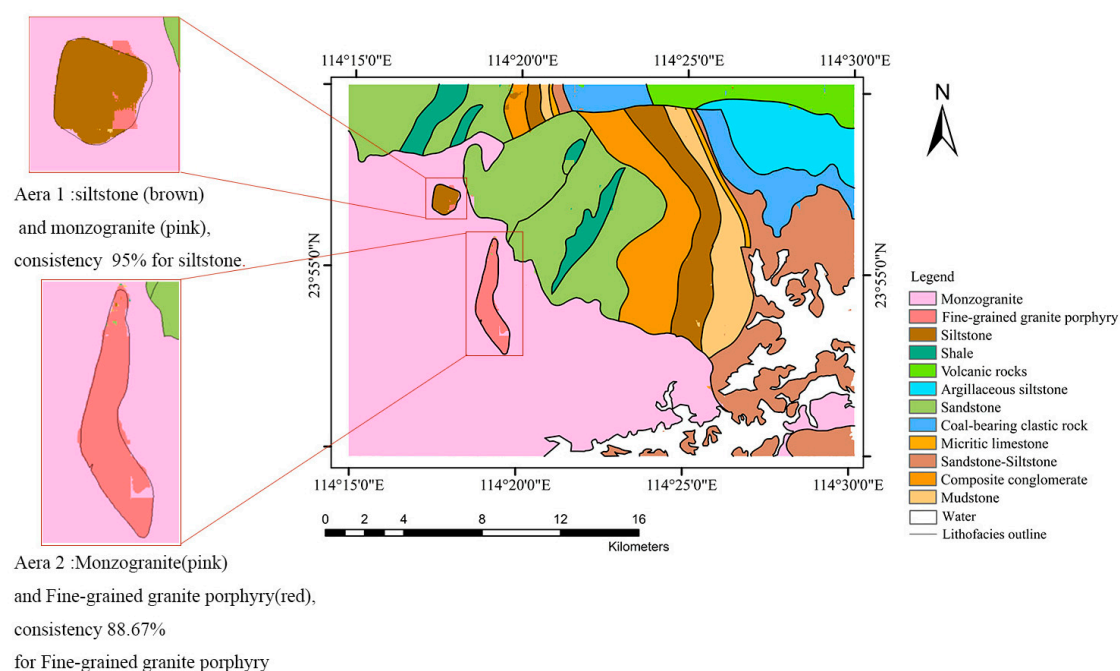


Figure 9. Model prediction lithological mapping (consistency refers to the pixel matching rate between the model's predicted results and the ground truth geological map).

Table 2 lists the quantitative evaluation metrics, including the IoU and F1 score, for the 12 lithological classes and two additional classes, mixed lithology and background. The results demonstrate that both individual lithologies and background classes achieved high IoU and F1 scores, indicating that the proposed model effectively distinguished between vegetation cover and lithological features. The model performed well in identifying sedimentary and granitic rocks.

Table 2. Prediction evaluation index results of 13 categories.

Class	IoU	Recall	Precision	F1-Score
Background	93.14%	97.16%	95.74%	96.44%
Monzogranite	97.69%	98.63%	99.03%	98.83%
Fine-grained granite porphyry	79.65%	91.12%	86.35%	88.67%
Siltstone	91.23%	96.05%	94.79%	95.42%
Shale	85.77%	89.38%	95.51%	92.34%
Volcanic rocks	94.19%	98.57%	95.50%	97.01%
Argillaceous siltstone	96.65%	97.24%	99.37%	98.29%
Sandstone	95.46%	98.24%	97.12%	97.68%
Coal-bearing clastic rock	92.83%	97.04%	95.53%	96.28%
Micritic limestone	71.68%	80.44%	86.80%	83.50%
Sandstone-Siltstone	92.02%	96.20%	95.49%	95.84%
Composite conglomerate	96.66%	97.28%	99.28%	98.27%
Mudstone	88.72%	92.75%	95.33%	94.02%

However, the classification performance for certain lithologies remained suboptimal. For instance, micritic limestone achieved an IoU of only 71.68% and an F1 score of 83.50%, which were relatively low compared with those of the other classes. This result suggests that further enhancement is required to improve the model's ability to distinguish lithologies with subtle spectral and textural characteristics.

4.4. Model Comparison Experiments

We conducted a series of comparative experiments using several common DL approaches to evaluate the performance of the proposed method (Figure 1). These include (1) a deep attention-based network proposed by Li et al. [46], (2) a multisource data feature fusion-based DL method proposed by Wang et al. [47], and (3) an improved U-Net model for lithological classification introduced by Brandmeier and Chen [48]. The single-branch ERFNet model proposed by Romera et al. [36] was adopted as the baseline for comparison to further validate the effectiveness of the dual-branch design.

The confusion matrices, OA metrics, and per-class F1-score statistics for all methods are presented in Figure 10, and Table 3. The proposed method achieved significantly superior performance compared with competing models (the OA result surpassed that of the best-performing improved U-Net model by 2.18%, the MIoU improved by 6.19% over the same model, and the MPA was 4.99% higher than that of the best-performing single-branch ERFNet baseline).

Table 3. Comparative experimental evaluation index results.

Method	Accuracy	MIoU	MPA
Single-branch ERFNet [36]	93.50%	81.77%	89.63%
Improved U-Net [48]	95.06%	84.24%	89.55%
Multisource data feature fusion [43]	86.46%	62.30%	72.44%
Attention-based deep network [46]	89.21%	71.33%	81.20%
TSFDNet (proposed model)	97.24%	90.43%	94.62%

Regarding the per-class classification capability, the proposed method outperformed all comparative models in terms of accuracy and F1 score for most lithologies, with the exception of mudstone and micritic limestone, where the performance gains were less prominent.

To compare the recognition stability of the models, we randomly split the datasets into 8:2 ratios, repeated the experiments five times, and recorded each model's F1 scores for all lithologies. We calculated the mean and standard deviation of the F1 scores per lithology for each model and plotted a bar chart with error bars (Figure 11). This chart reflects the F1-score fluctuation; smaller fluctuations indicate higher recognition stability for that lithology. The TSFDNet showed superior stability.

To further demonstrate the lithological identification capability of the five models, their respective lithological classification maps are presented (Figure 12). The results show that our model yielded more accurate class assignments, exhibited clearer lithological boundaries, and produced spatial extents that were more consistent with the ground truth. However, minor misclassifications still occurred.

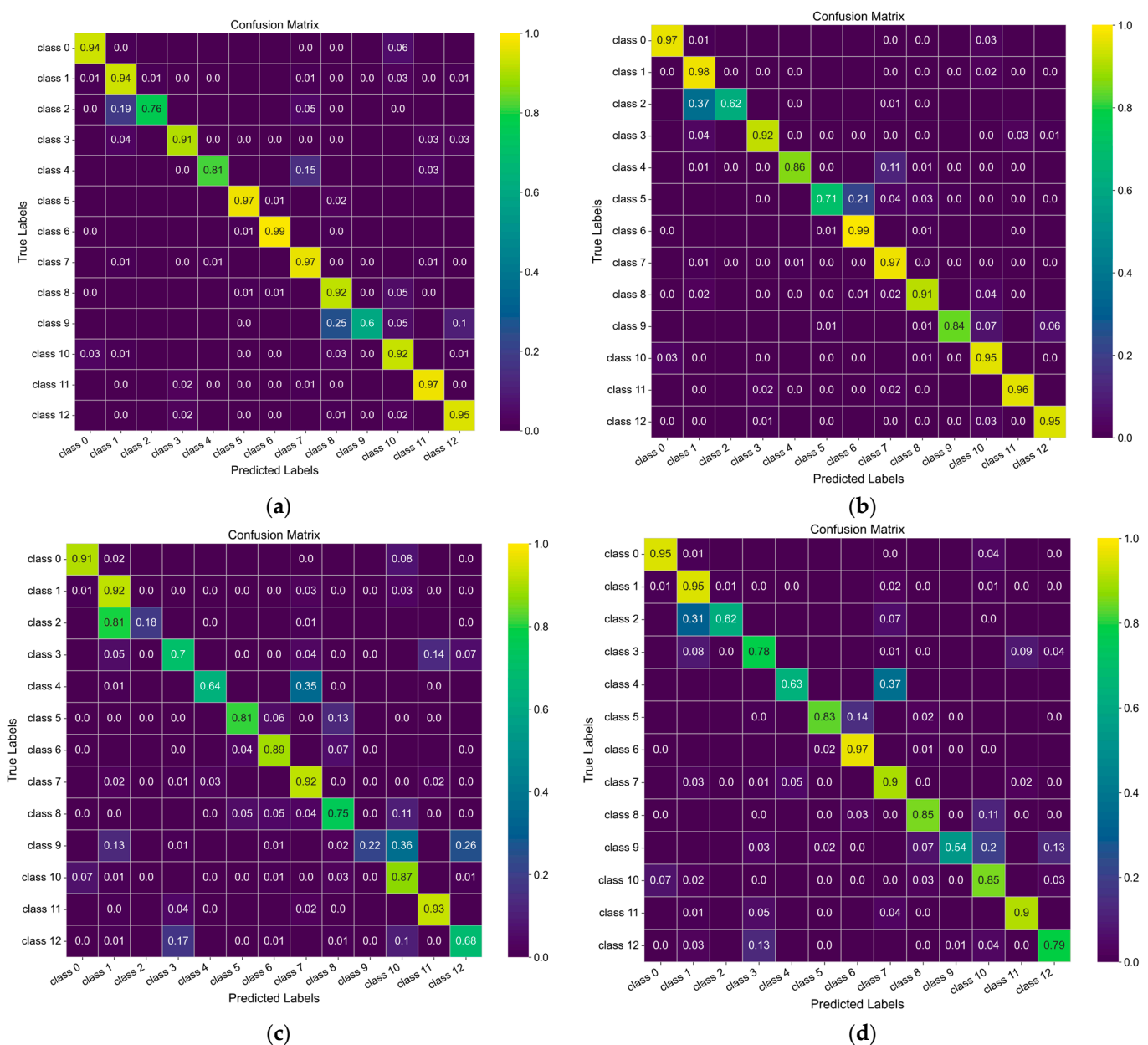


Figure 10. Cont.

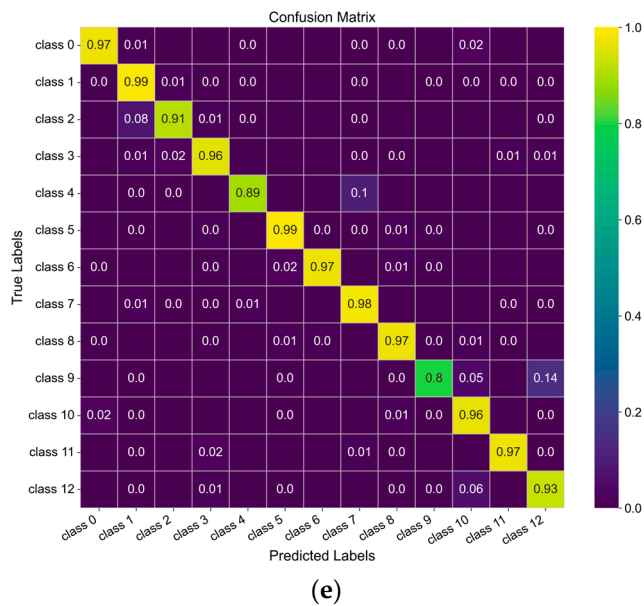


Figure 10. Confusion matrix of five methods: (a) Single-branch ERFNet [32], (b) improved U-Net [44], (c) Multisource data feature fusion [43], (d) attention-based deep network [42], (e) TSFDNet (proposed model). class0: water, class1: monzogranite, class2: fine-grained granite porphyry, class3: siltstone, class4: shale, class5: volcanic rocks, class6: argillaceous siltstone, class7: sandstone, class8: coal-bearing clastic rock, class9: micritic limestone, class10: sandstone-siltstone, class11: composite conglomerate, class12: mudstone.

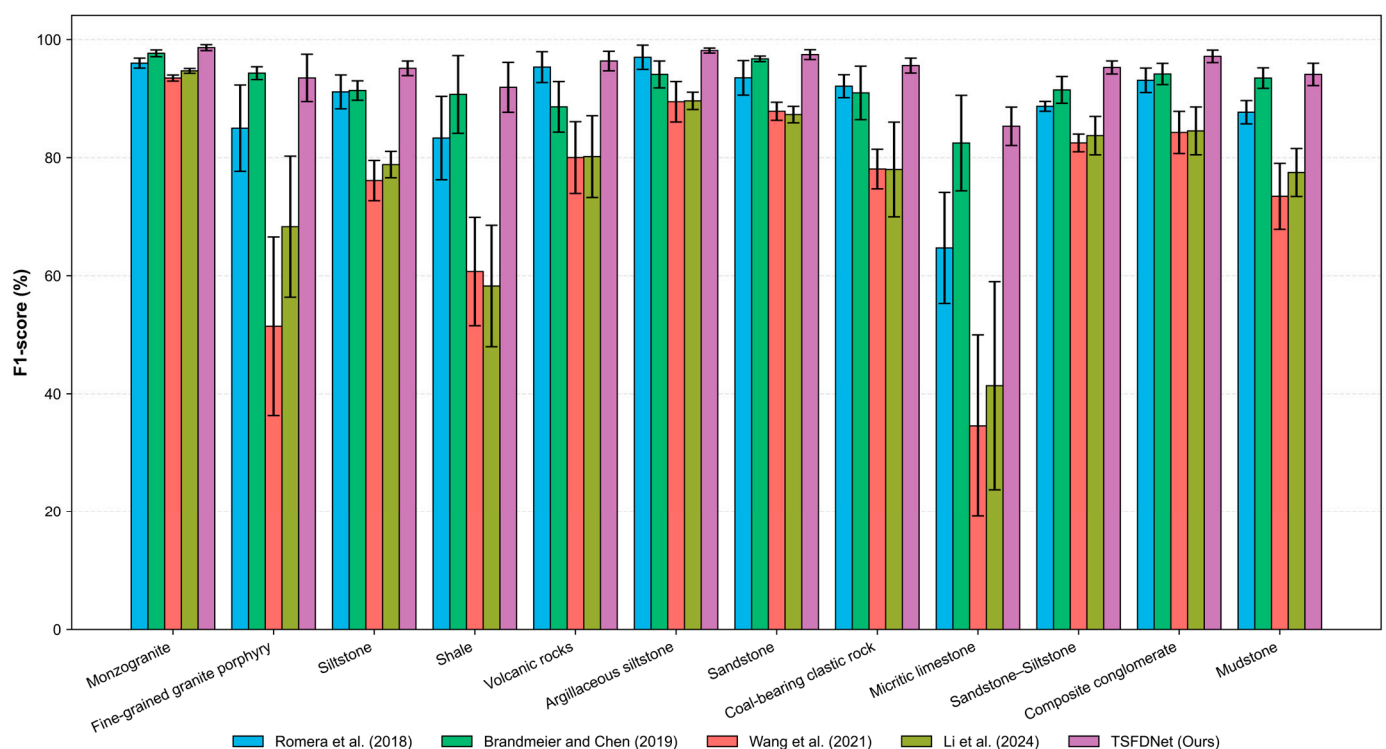


Figure 11. F1-score comparison error bar chart [36,43,46,48].

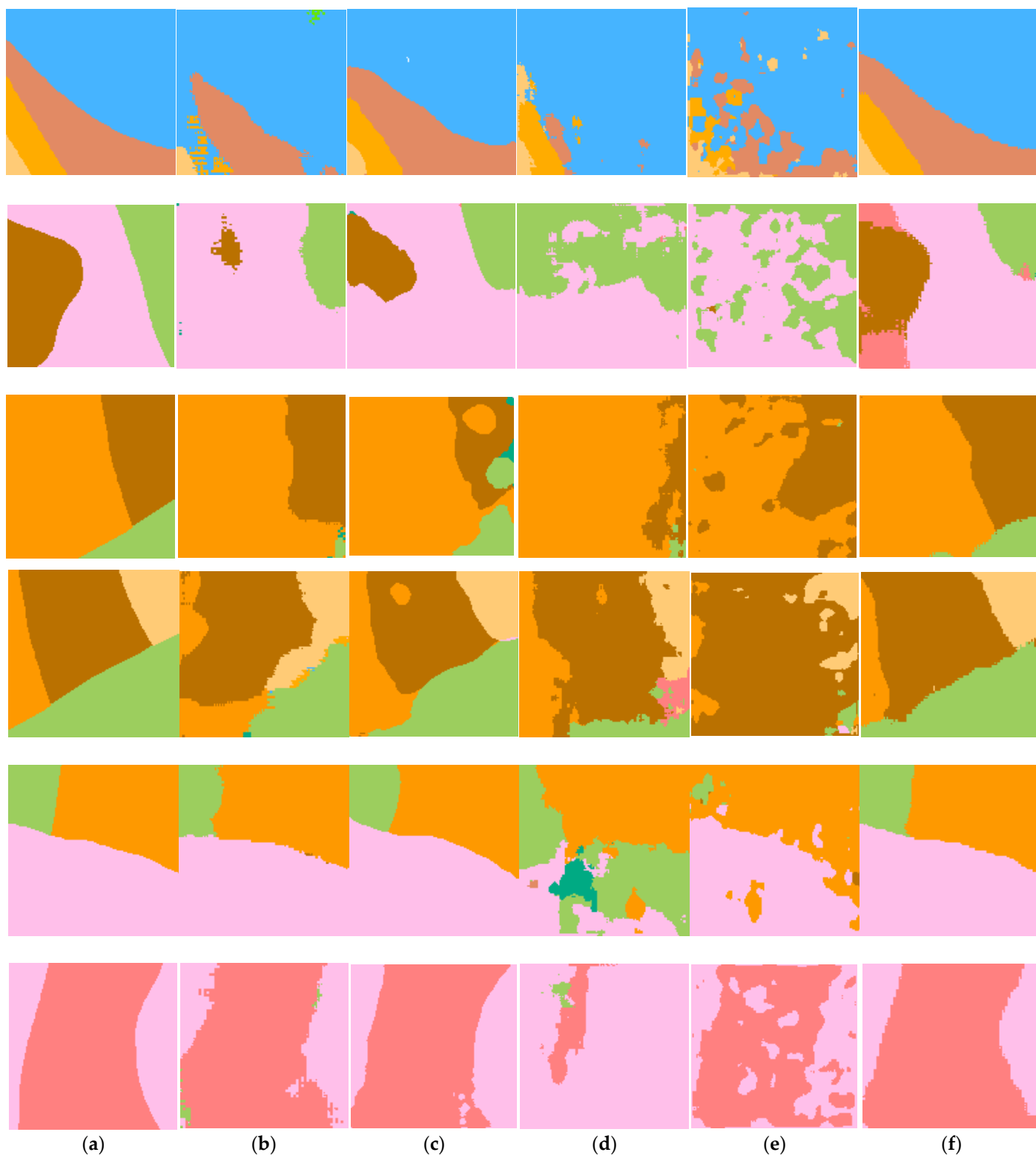


Figure 12. Comparison of lithology detail prediction results of five methods: (a) Labels, (b) single-branch ERFNet [32], (c) improved U-Net [44], (d) multi-source data feature fusion [43], (e) attention-based deep network [42], (f) TSFDNet (proposed model). The category information corresponding to the colors is shown in Figure 9.

5. Discussion

In recent years, numerous studies have indicated that for lithological classification tasks, especially under vegetation cover interference, multimodal remote sensing data

outperform single-source data [30,49–51]. However, these methods often rely merely on simple concatenation or stacking strategies [52,53], thereby overlooking the intrinsic complementarity among multi-source data. To address this issue, we propose TSFDNet. Experiments show that this framework boosted the overall classification accuracy by 10.78% over the baseline, and achieves higher spatial consistency across all 12 lithological classes.

Our work is mainly developed along the following aspects: (1) A dual-branch architecture is employed to decouple modality-specific features. In the hyperspectral branch, an SSFM module is introduced to enable deep extraction of mineral spectral characteristics from Gaofen-5 data. In the microwave branch, a TFEM module is incorporated to capture deep indirect stratigraphic texture features from Sentinel-1 C-band observations. (2) An FFEM module is designed to achieve adaptive fusion of spectral and microwave features.

In terms of model performance, we demonstrate that TSFDNet serves as an effective deep learning framework for lithology identification in regions with dense vegetation cover. The dual-branch architecture of TSFDNet outperforms conventional single-branch designs. Experimental results show that our method achieves superior boundary preservation, fine-grained class discrimination, and model stability (Figures 9 and 12). Within TSFDNet, the SSFM module captures long-range spectral dependencies and local spatial structures through a three-branch spectral–spatial modeling strategy. And the TFEM module enhances the representation of microwave-derived textures by explicitly modeling dominant texture orientations and critical spatial regions. Compared with dual-branch networks that rely solely on single-path attention mechanisms or simple convolutional stacking—often leading to misclassification within lithologically homogeneous units [46]—TSFDNet shows markedly improved stability and discriminative ability, especially for lithologies with similar spectral or textural expressions such as sandstone and siltstone. Moreover, the FFEM module adaptively modulates the contribution of hyperspectral and microwave features based on modality specificity and an EMA-guided strategy, strengthening key region responses and improving lithology classification accuracy. Unlike early-stage feature concatenation approaches [36,48], the inclusion of FFEM enables TSFDNet to maintain spatial continuity and consistency along lithological boundaries while reducing noise within homogeneous units. Although some studies perform early-stage feature-level fusion, such as GS fusion [47], these methods do not explicitly differentiate between hyperspectral and microwave modalities, often introducing redundant features and noise at shallow levels. This limitation is particularly pronounced in areas with dense vegetation cover, where the complementary advantages of multi-source data cannot be fully exploited, ultimately constraining classification accuracy.

In addition to high lithological recognition accuracy, the proposed model also exhibits excellent stability (Figure 11). After multiple repeated experiments, TSFDNet showed smaller F1-score fluctuations than the other models for sedimentary rocks with similar spectral curves (e.g., sandstone–siltstone, siltstone, and sandstone), which were stable at 98.2%, 94%, and 98%, respectively. However, the stability of TSFDNet in granite identification still requires further improvement. Overall, These results demonstrate that our model has strong stability in distinguishing similar lithologies and handling small-sample cases. TSFDNet demonstrated high consistency between the predicted lithological map and the actual distribution.

It is worth noting that, in selecting the data source for texture characterization, we intentionally opted for the C-band rather than longer-wavelength L- or P-band data, despite their stronger penetration capability. This choice is motivated by several considerations. First, although C-band backscattering in vegetated areas is generally dominated by canopy responses, the vegetation in our study area is characterized by relatively thin trunks and limited canopy-to-canopy occlusion. Consequently, the attenuation effect is less pronounced

than in densely forested regions. Second, the texture features captured by our model should be understood as indirect surface textures rather than direct geological textures. Different lithologies exhibit varying soil thicknesses and weathering products [54,55], which in turn induce subtle differences in canopy roughness, scattering configurations, local slope patterns, and microscale surface roughness even under the same vegetation type [19,52,56]. These secondary variations can still be expressed in the C-band backscatter signal. Furthermore, the feasibility of using C-band data for lithology identification in vegetated regions has been validated in previous studies [27,57–59]. Our experimental results align with these findings and further confirm that C-band-derived textures are sufficiently sensitive to lithology-induced surface variations in the study area.

Despite the excellent performance of the TSFDNet, minor misclassifications were observed. For example, micritic limestone has a relatively low accuracy owing to its narrow-banded distribution and high spectral-textural similarity to the surrounding sandstone–siltstone and mudstone, and their spectral curves differed mainly at 800–1300 nm, while showing high similarity elsewhere. In contrast, the monzonitic granite achieved a significantly higher accuracy owing to the abundance of samples. These results confirm that class imbalance and intraclass spectral similarity remain the key challenges affecting model performance. Therefore, future research should explore the integration of thermal infrared or geophysical datasets to expand the feature dimensionality of the model [60,61]. Furthermore, techniques including data augmentation, resampling strategies, and class weight adjustment were investigated to mitigate the adverse impacts induced by class imbalance and interclass feature ambiguity, thereby further enhancing the robustness and generalization capability of the proposed framework.

6. Conclusions

Vegetation coverage degrades the lithological classification accuracy of traditional remote sensing methods. Hyperspectral remote sensing provides fine-grained spatial-spectral information, whereas microwave remote sensing offers strong penetration and prominent texture expression. To leverage the complementary characteristics of these two types of data, we developed TSFDNet for lithology classification in areas with vegetation cover and reached the following key conclusions:

1. The encoder–decoder-based deep learning architecture exhibits a strong classification capability in complex environments (e.g., vegetation-covered areas).
2. The dual-branch structure significantly enhances multisource data representational capability. By designing independent extraction paths for indirect lithological texture features, which include vegetation structural information, and spectral features, TSFDNet reduces feature interference and information loss, and it improves the OA and MIoU by 2.1% and 3.4%, respectively, compared with single-branch fusion networks.
3. The dedicated feature extraction and fusion modules substantially improve the classification performance. Three synergistic modules were designed to optimize the process of feature learning and integration, enabling the model to achieve an overall classification accuracy of 97.24%, with excellent robustness.

Ultimately, the findings of this study advance multisource remote sensing fusion for lithological mapping. They offer a scalable solution to address the longstanding challenges in vegetation-covered and geologically complex areas. This work provides a practical and scalable framework for leveraging complementary multisource remote sensing data, paving the way for more accurate and efficient lithological mapping in challenging, vegetation-covered regions.

Author Contributions: Conceptualization, Y.X. and Z.Z.; methodology, Z.Z.; software, Z.Z.; validation, Z.Z., Y.X. and J.C.; formal analysis, Z.Z.; investigation, Y.X.; resources, J.C.; data curation, J.C.; writing—original draft preparation, Z.Z.; writing—review and editing, Y.X.; visualization, Z.Z.; supervision, Y.X.; project administration, Y.X.; funding acquisition, Y.X. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Key R&D Program of China under Grant 2023YFC2906404 and the Research Project of China Water Resources Pearl River Planning Surveying & Designing Co., Ltd. under Grant 2023KY01.

Data Availability Statement: The Sentinel-1 data are available at <https://www.usgs.gov/labs/spectroscopylab/science/spectral-library> (accessed on 20 December 2024). GF-5 and geological maps are not publicly available because these data were also used in other experimental studies and cannot be made public.

Acknowledgments: This work was supported by the National Key R&D Program of China under Grant 2023YFC2906404 and the Research Project of China Water Resources Pearl River Planning Surveying & Designing Co., Ltd. under Grant 2023KY01.

Conflicts of Interest: The funding sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results. The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this paper:

EMA	Efficient multi-scale attention
ERFNet	Efficient residual factorized ConvNet
FFEM	Feature fusion extraction module
FN	False negative
FP	False positive
GAP	Global average pooling
GF-5	Gaofen-5
GLCM	Gray-level co-occurrence matrix
GRD	Ground range detected
mIoU	Mean Intersection over Union
MPA	Mean pixel accuracy
Non-bt-1D	Non-bottleneck-1D
OA	Overall classification accuracy
PCA	Principal component analysis
SAR	Synthetic aperture radar
SNAP	Sentinel application platform
SSFM	Spectral-spatial feature module
TFEM	Texture feature extraction module
TN	True negative
TP	True positive
TSFDNet	Dual-branch texture and spectral feature network

References

1. 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](#)]
2. Wang, Z.; Zuo, R. Intelligent Lithological Mapping: Challenges and Future Prospective. *Nat. Resour. Res.* **2025**, 1–34. [[CrossRef](#)]
3. Grebby, S.; Cunningham, D.; Naden, J.; Tansey, K. Lithological Mapping of the Troodos Ophiolite, Cyprus, Using Airborne LiDAR Topographic Data. *Remote Sens. Environ.* **2010**, *114*, 713–724. [[CrossRef](#)]
4. Gad, S.; Kusky, T. Lithological Mapping in the Eastern Desert of Egypt, the Barramiya Area, Using Landsat Thematic Mapper (TM). *J. Afr. Earth Sci.* **2006**, *44*, 196–202. [[CrossRef](#)]

5. Pereira, J.; Pereira, A.; Gil, A.; Mantas, V.M. Lithology Mapping with Satellite Images, Fieldwork-Based Spectral Data, and Machine Learning Algorithms: The Case Study of Beiras Group (Central Portugal). *Catena* **2023**, *220*, 106653. [\[CrossRef\]](#)
6. Rezaei, A.; Hassani, H.; Moarefvand, P.; Golmohammadi, A. Lithological Mapping in Sangan Region in Northeast Iran Using ASTER Satellite Data and Image Processing Methods. *Geol. Ecol. Landsc.* **2020**, *4*, 59–70. [\[CrossRef\]](#)
7. Ali, E.-O.M.; Abdelkader, E.G. A Review on Advancements in Lithological Mapping Utilizing Machine Learning Algorithms and Remote Sensing Data. *Heliyon* **2023**, *9*, e20168. [\[CrossRef\]](#)
8. Peyghambari, S.; Zhang, Y. Hyperspectral Remote Sensing in Lithological Mapping, Mineral Exploration, and Environmental Geology: An Updated Review. *J. Appl. Remote Sens.* **2021**, *15*, 031501. [\[CrossRef\]](#)
9. Chandan, K.; Snehamoy, C.; Thomas, O.; Arindam, G. 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.* **2019**, *86*, 102006.
10. Eldosouky, A.M.; Othman, A.; Saada, S.A.; Zamzam, S. A New Vector for Mapping Gold Mineralization Potential and Proposed Pathways in Highly Weathered Basement Rocks Using Multispectral, Radar, and Magnetic Data in Random Forest Algorithm. *Nat. Resour. Res.* **2023**, *33*, 23–50. [\[CrossRef\]](#)
11. Kewang, D.; Huijie, Z.; Na, L.; Wei, W. Identification of Minerals in Hyperspectral Imagery Based on the Attenuation Spectral Absorption Index Vector Using a Multilayer Perceptron. *Remote Sens. Lett.* **2021**, *12*, 449–458.
12. 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. [\[CrossRef\]](#)
13. Molan, Y.E.; Refahi, D.; Tarashti, A.H. Mineral Mapping in the Maherabad Area, Eastern Iran, Using the HyMap Remote Sensing Data. *Int. J. Appl. Earth Obs. Geoinf.* **2014**, *27*, 117–127. [\[CrossRef\]](#)
14. Zadeh, M.H.; Tangestani, M.H.; Roldan, F.V.; Yusta, I. Sub-Pixel Mineral Mapping of a Porphyry Copper Belt Using EO-1 Hyperion Data. *Adv. Space Res.* **2014**, *53*, 440–451. [\[CrossRef\]](#)
15. Li, N.; Huang, X.; Zhao, H.; Qiu, X.; Geng, R.; Jia, X.; Wang, D. Multiparameter Optimization for Mineral Mapping Using Hyperspectral Imagery. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2018**, *11*, 1348–1357. [\[CrossRef\]](#)
16. Ni, L.; Xu, H.; Zhou, X. Mineral Identification and Mapping by Synthesis of Hyperspectral VNIR/SWIR and Multispectral TIR Remotely Sensed Data with Different Classifiers. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *13*, 3155–3163. [\[CrossRef\]](#)
17. Chen, Y.; Wang, Y.; Zhang, F.; Dong, Y.; Song, Z.; Liu, G. Remote Sensing for Lithology Mapping in Vegetation-Covered Regions: Methods, Challenges, and Opportunities. *Minerals* **2023**, *13*, 1153. [\[CrossRef\]](#)
18. Haest, M.; Cudahy, T.; Rodger, A.; Laukamp, C.; Martens, E.; Caccetta, M. Unmixing the Effects of Vegetation in Airborne Hyperspectral Mineral Maps over the Rocklea Dome Iron-Rich Palaeochannel System (Western Australia). *Remote Sens. Environ.* **2013**, *129*, 17–31. [\[CrossRef\]](#)
19. do Amaral, C.H.; de Almeida, T.I.R.; de Souza Filho, C.R.; Roberts, D.A.; Fraser, S.J.; Alves, M.N.; Botelho, M. Characterization of Indicator Tree Species in Neotropical Environments and Implications for Geological Mapping. *Remote Sens. Environ.* **2018**, *216*, 385–400. [\[CrossRef\]](#)
20. Okada, N.; Maekawa, Y.; Owada, N.; Haga, K.; Shibayama, A.; Kawamura, Y. Automated Identification of Mineral Types and Grain Size Using Hyperspectral Imaging and Deep Learning for Mineral Processing. *Minerals* **2020**, *10*, 809. [\[CrossRef\]](#)
21. Wang, S.; Zhou, Q. Multi-Source Fusion Enhanced Feature Segmentation in Remote Sensing Imagery. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2024**, *10*, 395–401. [\[CrossRef\]](#)
22. Rowan, L.C.; Crowley, J.K.; Schmidt, R.G.; Ager, C.M.; Mars, J.C. Mapping Hydrothermally Altered Rocks by Analyzing Hyperspectral Image (AVIRIS) Data of Forested Areas in the Southeastern United States. *J. Geochem. Explor.* **2000**, *68*, 145–166. [\[CrossRef\]](#)
23. Meroni, M.; d’Andrimont, R.; Vrieling, A.; Fasbender, D.; Lemoine, G.; Rembold, F.; Seguini, L.; Verhegghen, A. Comparing Land Surface Phenology of Major European Crops as Derived from SAR and Multispectral Data of Sentinel-1 and-2. *Remote Sens. Environ.* **2021**, *253*, 112232. [\[CrossRef\]](#) [\[PubMed\]](#)
24. Mastrorosa, S.; Crespi, M.; Congedo, L.; Munafò, M. Land Consumption Classification Using Sentinel 1 Data: A Systematic Review. *Land* **2023**, *12*, 932. [\[CrossRef\]](#)
25. Guo, S.; Yang, C.; Han, L.; Feng, Y.; Zhao, J. Lithology Classification of Igneous Rocks Using C-Band and L-Band Dual-Polarization SAR Data. *Open Geosci.* **2023**, *15*, 20220465. [\[CrossRef\]](#)
26. Fu, D.; Su, C.; Wang, W.; Yuan, R. Deep Learning Based Lithology Classification of Drill Core Images. *PLoS ONE* **2022**, *17*, e0270826. [\[CrossRef\]](#)
27. Lu, Y.; Yang, C.; Meng, Z. Lithology Discrimination Using Sentinel-1 Dual-Pol Data and SRTM Data. *Remote Sens.* **2021**, *13*, 1280. [\[CrossRef\]](#)
28. 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. [\[CrossRef\]](#)

29. Li, R.; Zhou, M.; Zhang, D.; Yan, Y.; Huo, Q. A Survey of Multi-Source Image Fusion. *Multimed. Tools Appl.* **2024**, *83*, 18573–18605. [\[CrossRef\]](#)
30. Zhang, J. Multi-Source Remote Sensing Data Fusion: Status and Trends. *Int. J. Image Data Fusion* **2010**, *1*, 5–24. [\[CrossRef\]](#)
31. Zuo, W.; Zeng, X.; Gao, X.; Zhang, Z.; Liu, D.; Li, C. Machine Learning Fusion Multi-Source Data Features for Classification Prediction of Lunar Surface Geological Units. *Remote Sens.* **2022**, *14*, 5075. [\[CrossRef\]](#)
32. 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. [\[CrossRef\]](#)
33. Lu, J.; Li, L.; Wang, J.; Han, L.; Xia, Z.; He, H.; Bai, Z. MSIMRS: Multi-Scale Superpixel Segmentation Integrating Multi-Source Remote Sensing Data for Lithology Identification in Semi-Arid Area. *Remote Sens.* **2025**, *17*, 387. [\[CrossRef\]](#)
34. Han, W.; Li, J.; Wang, S.; Zhang, X.; Dong, Y.; Fan, R.; Zhang, X.; Wang, L. Geological Remote Sensing Interpretation Using Deep Learning Feature and an Adaptive Multisource Data Fusion Network. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–14. [\[CrossRef\]](#)
35. Xu, F.; Wang, B. Debris Flow Susceptibility Mapping in Mountainous Area Based on Multi-Source Data Fusion and CNN Model—Taking Nujiang Prefecture, China as an Example. *Int. J. Digit. Earth* **2022**, *15*, 1966–1988. [\[CrossRef\]](#)
36. Romera, E.; Alvarez, J.M.; Bergasa, L.M.; Arroyo, R. ERFNet: Efficient Residual Factorized ConvNet for Real-Time Semantic Segmentation. *IEEE Trans. Intell. Transp. Syst.* **2018**, *19*, 263–272. [\[CrossRef\]](#)
37. Guo, S.; Yang, C.; He, R.; Li, Y. Improvement of Lithological Mapping Using Discrete Wavelet Transformation from Sentinel-1 SAR Data. *Remote Sens.* **2022**, *14*, 5824. [\[CrossRef\]](#)
38. El Rai, M.C.; Darweesh, M.; Far, A.B.; Gawanmeh, A. Multi-Source Attention-Based Fusion for Segmentation of Natural Disasters. *IEEE Geosci. Remote Sens. Lett.* **2024**, *21*, 1–5. [\[CrossRef\]](#)
39. Shang, R.; Peng, P.; Shang, F.; Jiao, L.; Shen, Y.; Stolkin, R. Semantic Segmentation for SAR Image Based on Texture Complexity Analysis and Key Superpixels. *Remote Sens.* **2020**, *12*, 2141. [\[CrossRef\]](#)
40. Li, X.; Zhang, G.; Cui, H.; Hou, S.; Wang, S.; Li, X.; Chen, Y.; Li, Z.; Zhang, L. MCANet: A Joint Semantic Segmentation Framework of Optical and SAR Images for Land Use Classification. *Int. J. Appl. Earth Obs. Geoinf.* **2022**, *106*, 102638. [\[CrossRef\]](#)
41. Fang, Y.; Sun, L.; Zheng, Y.; Wu, Z. Deformable Convolution-Enhanced Hierarchical Transformer with Spectral-Spatial Cluster Attention for Hyperspectral Image Classification. *IEEE Trans. Image Process.* **2025**, *34*, 701–716. [\[CrossRef\]](#) [\[PubMed\]](#)
42. He, M.; Li, B.; Chen, H. Multi-Scale 3D Deep Convolutional Neural Network for Hyperspectral Image Classification. In Proceedings of the 2017 IEEE International Conference on Image Processing (ICIP), Beijing, China, 17–20 September 2017; pp. 3904–3908.
43. Liu, C.; Sun, Y.; Xu, Y.; Sun, Z.; Zhang, X.; Lei, L.; Kuang, G. A Review of Optical and SAR Image Deep Feature Fusion in Semantic Segmentation. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2024**, *17*, 12910–12930. [\[CrossRef\]](#)
44. Ouyang, D.; He, S.; Zhang, G.; Luo, M.; Guo, H.; Zhan, J.; Huang, Z. Efficient Multi-Scale Attention Module with Cross-Spatial Learning. In Proceedings of the ICASSP 2023—2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Rhodes Island, Greece, 4–10 June 2023; pp. 1–5.
45. Wang, S.; Huang, X.; Han, W.; Li, J.; Zhang, X.; Wang, L. Lithological Mapping of Geological Remote Sensing via Adversarial Semi-Supervised Segmentation Network. *Int. J. Appl. Earth Obs. Geoinf.* **2023**, *125*, 103536. [\[CrossRef\]](#)
46. Li, D.; Wang, J.; Zhou, K.; Bi, J.; Zhang, Q.; Wang, W.; Qu, G.; Li, C.; Qiu, H.; Liao, T.; et al. Application of Hybrid Attention Mechanisms in Lithological Classification with Multisource Data: A Case Study from the Altay Orogenic Belt. *Remote Sens.* **2024**, *16*, 3958. [\[CrossRef\]](#)
47. Wang, Z.; Zuo, R.; Liu, H. Lithological Mapping Based on Fully Convolutional Network and Multi-Source Geological Data. *Remote Sens.* **2021**, *13*, 4860. [\[CrossRef\]](#)
48. Brandmeier, M.; Chen, Y. Lithological Classification Using Multi-Sensor Data and Convolutional Neural Networks. *ISPRS—Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2019**, *XLII-2/W16*, 55–59. [\[CrossRef\]](#)
49. Bachri, I.; Hakdaoui, M.; Raji, M.; Teodoro, A.C.; Benbouziane, A. 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. Geo-Inf.* **2019**, *8*, 248. [\[CrossRef\]](#)
50. Shayeganpour, S.; Tangestani, M.H.; Gorsevski, P.V. Machine Learning and Multi-Sensor Data Fusion for Mapping Lithology: A Case Study of Kowli-Kosh Area, SW Iran. *Adv. Space Res.* **2021**, *68*, 3992–4015. [\[CrossRef\]](#)
51. Kumar, C.; Chatterjee, S.; Oommen, T.; Guha, A.; Mukherjee, A. Multi-Sensor Datasets-Based Optimal Integration of Spectral, Textural, and Morphological Characteristics of Rocks for Lithological Classification Using Machine Learning Models. *Geocarto Int.* **2022**, *37*, 6004–6032. [\[CrossRef\]](#)
52. Grebby, S.; Naden, J.; Cunningham, D.; Tansey, K. Integrating Airborne Multispectral Imagery and Airborne LiDAR Data for Enhanced Lithological Mapping in Vegetated Terrain. *Remote Sens. Environ.* **2011**, *115*, 214–226. [\[CrossRef\]](#)
53. Ge, W.; Cheng, Q.; Tang, Y.; Jing, L.; Gao, C. Lithological Classification Using Sentinel-2A Data in the Shibanzhong Ophiolite Complex in Inner Mongolia, China. *Remote Sens.* **2018**, *10*, 638. [\[CrossRef\]](#)

54. Schwinning, S. The Ecohydrology of Roots in Rocks. Ecohydrology: Ecosystems, land and water process interactions. *Ecohydrology* **2010**, *3*, 238–245. [[CrossRef](#)]
55. Klos, P.Z.; Goulden, M.L.; Riebe, C.S.; Tague, C.L.; O'Geen, A.T.; Flinchum, B.A.; Safeeq, M.; Conklin, M.H.; Hart, S.C.; Berhe, A.A.; et al. Subsurface Plant-Accessible Water in Mountain Ecosystems with a Mediterranean Climate. *Wiley Interdiscip. Rev. Water* **2018**, *5*, e1277. [[CrossRef](#)]
56. Vitousek, P.; Asner, G.P.; Chadwick, O.A.; Hotchkiss, S. Landscape-Level Variation in Forest Structure and Biogeochemistry across a Substrate Age Gradient in Hawaii. *Ecology* **2009**, *90*, 3074–3086. [[CrossRef](#)] [[PubMed](#)]
57. Chen, Y.; Liu, G.; Song, Z.; Li, M.; Wang, M.; Wang, S. Lithological Mapping in High-Vegetation Areas Using Sentinel-2, Sentinel-1, and Digital Elevation Models. *Sensors* **2025**, *25*, 2136. [[CrossRef](#)]
58. Manap, H.S.; San, B.T. Data Integration for Lithological Mapping Using Machine Learning Algorithms. *Earth Sci. Inform.* **2022**, *15*, 1841–1859. [[CrossRef](#)]
59. Desoky, H.A.; Abd El-Dayem, M.; Hegab, M.A.E.-R. A Comparative Analysis to Assess the Efficiency of Lineament Extraction Utilizing Satellite Imagery from Landsat-8, Sentinel-2B, and Sentinel-1A: A Case Study around Suez Canal Zone, Egypt. *Remote Sens. Appl. Soc. Environ.* **2024**, *36*, 101312. [[CrossRef](#)]
60. Fahmy, W.; El-Desoky, H.M.; Elyaseer, M.H.; Ayonta Kenne, P.; Shirazi, A.; Hezarkhani, A.; Shirazy, A.; El-Awny, H.; Abdel-Rahman, A.M.; Khalil, A.E.; et al. Remote Sensing, Petrological and Geochemical Data for Lithological Mapping in Wadi Kid, Southeast Sinai, Egypt. *Minerals* **2023**, *13*, 1160. [[CrossRef](#)]
61. Slavinski, H.; Morris, B.; Ugalde, H.; Spicer, B.; Skulski, T.; Rogers, N. Integration of Lithological, Geophysical, and Remote Sensing Information: A Basis for Remote Predictive Geological Mapping of the Baie Verte Peninsula, Newfoundland. *Can. J. Remote Sens.* **2010**, *36*, 99–118. [[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.