remote sensing

MDPI

Article

Mapping Lithology with Hybrid Attention Mechanism–Long Short-Term Memory: A Hybrid Neural Network Approach Using Remote Sensing and Geophysical Data

Michael Appiah-Twum [1,2,*], Wenbo Xu [1,2] and Emmanuel Daanoba Sunkari [3,4]

[1] School of Resources and Environment, University of Electronic Science and Technology of China, Chengdu 611731, China; xuwenbo@uestc.edu.cn
[2] Yangtze Delta Region Institute (Huzhou), University of Electronic Science and Technology of China, Huzhou 313001, China
[3] Department of Chemical Sciences, Faculty of Science, University of Johannesburg, Auckland Park, Johannesburg P.O. Box 524, South Africa
[4] Department of Geological Engineering, Faculty of Geosciences, University of Mines and Technology, Tarkwa P.O. Box 237, Ghana; edsunkari@umat.edu.gh
* Correspondence: michaelappiah-twum@hotmail.com or 202014070102@std.uestc.edu.cn

Abstract: Remote sensing provides an efficient roadmap in geological analysis and interpretation. However, some challenges arise when remote sensing techniques are integrated with machine learning in geological surveys. Factors including irregular spatial distribution, sample imbalance, interclass resemblances, regolith, and geochemical similarities impede geological feature diagnosis, interpretation, and identification across varied remote sensing datasets. To address these limitations, a hybrid-attention-integrated long short-term memory (LSTM) network is employed to diagnose, interpret, and identify lithological feature representations in a remote sensing-based geological analysis using multisource data fusion. The experimental design integrates varied datasets including Sentinel-2A, Landsat-9, ASTER, ALOS PALSAR DEM, and Bouguer anomaly gravity data. The proposed model incorporates a hybrid attention mechanism (HAM) comprising channel and spatial attention submodules. HAM utilizes an adaptive technique that merges global-average-pooled features with max-pooled features, enhancing the model's accuracy in identifying lithological units. Additionally, a channel separation operation is employed to allot refined channel features into clusters based on channel attention maps along the channel dimension. The comprehensive analysis of results from comparative extensive experiments demonstrates HAM-LSTM's state-of-the-art performance, outperforming existing attention modules and attention-based models (ViT, SE-LSTM, and CBAM-LSTM). Comparing HAM-LSTM to baseline LSTM, the HAM module's integrated configurations equip the proposed model to better diagnose and identify lithological units, thereby increasing the accuracy by 3.69%.

Keywords: deep learning; hybrid attention mechanism; HAM-LSTM; lithological mapping; multisensor multispectral data; geophysical data

check for updates

Citation: 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. https://doi.org/10.3390/rs16234613

Academic Editor: Lenio Soares Galvao

Received: 21 September 2024
Revised: 10 November 2024
Accepted: 12 November 2024
Published: 9 December 2024

Copyright: © 2024 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) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Mapping lithology holds significant importance in diverse fields including rural–urban development, environmental management, infrastructural design, and resource evaluation [1–4]. Traditional data collection methods in geological studies are predominantly characterized by exhaustive field surveys and manual lithological mapping. However, the effectiveness of such approaches is threatened by challenges such as the remoteness and inaccessibility of target terranes, harsh weather and environmental conditions, operational cost, and security. To address these challenges, the emergence of remote sensing offers a cost-effective and efficient alternative for data collection, thereby facilitating geological studies [5–7].

In the field of geosciences, integrating remote sensing technology with deep learning (DL) to map lithology has gained popularity given the ability of DL models to effectively handle high data dimensionality for classification and prediction tasks. In particular, convolutional neural networks (CNNs) provide effective utility in remote sensing-based geological feature analysis due to their robust deep structures and convolutional operations [8–11]. In learning and extracting features from input data, CNNs have demonstrated resilience in handling spatial disparities and have proven capable of distinguishing features in a given dataset with remarkable precision [10,12].

Beyond CNNs, gated recurrent unit (GRU) [13] and other DL techniques including long short-term memory (LSTM) networks [14] have shown significant progress in classifying remote sensing imagery. For instance, a recurrent neural network (RNN)-based framework, modeled to classify hyperspectral imagery, distinguishes information groups by utilizing network reasoning to handle hyperspectral pixels as sequential data [15]. In tackling local and semantic feature extraction challenges, modeling a sequential spatial feature learning approach in an RNN architecture proves to be efficient [16]. In other studies [17], the integration of CNNs with RNNs to generate hybrid variants leverages the advantages of RNNs and CNNs, culminating in efficient convolutional feature extraction with subsequent classification by recurrent layers. Essentially, in the RNN-based framework, input data are configured in series and iteratively processed sequentially with a linear interconnection of all recurrent units. In the context of classifying remote sensing imagery, remotely sensed data are sequentially inferred by an RNN along the spectral dimensions of the data with a focus on learning the variability in the spectral sequences in the data to distinguish pixel classes [16,17].

The application of DL models in remote sensing-based geological interpretation has been carried out by several studies using varied datasets spanning hyperspectral, multispectral, and unmanned aerial vehicles (UAVs) [9,18–20]. Despite the significant contributions, the following challenges persist in the implementation of DL models in remote sensing-based geological interpretation:

1. *Single-sensor information capture*: Most of the existing sensors in remote sensing have unique configurations and thus capture information using their defined configurations. Thus, information captured by a single data source might not be sufficient for analysis. Exemplified by the limitation of using a single data source for the diagnosis of target lithological feature representations in intricate conditions, it, therefore, becomes necessary to integrate data from varied sources to improve data credibility and enhance target feature diagnosis.
2. *Data volume and computational complexity*: DL implementations in remote sensing-based geological interpretation mostly rely on 2D interpretations, which thereby requires extensive image segmentation and a bigger data volume for model training. Relative to individual pixel analysis in traditional image classification methods, predictions in CNN implementations are often dependent on the patches within a pixel's circumference. Challenges such as snags in the computation of optimal patch sizes arise in this approach, especially in domains where there are overlapping geological features. Such difficulties impede model performance as a result of the unnecessary computational complexities [7].
3. *Spectral and spatial feature learning of RNNs*: Notwithstanding the progress of RNNs in classifying remote sensing imagery, their implementation faces some concerns. Chiefly amongst such concerns is the limitation in comprehensively assessing spectral similarities within an input dataset caused by their sequential iterative processing of individual pixels along the spectral axis. Furthermore, the direct integration of the spatial context in feature learning, which is imperative for achieving appreciable accuracy in a classification task, is a challenge in traditional RNNs. Despite the efforts of bidirectional LSTM in integrating spectral and spatial feature learning [21], the resultant output of learned spectral and spatial features remains inadequate given

that features are learned separately, and the learned spatial and spectral features are thereby subsequently concatenated for the classification task.

In addressing the aforementioned challenges, this study proposes an alternative approach employing hybrid attention with an LSTM network to comprehensively assess the spectral interrelation amongst feature representations in multisource data for geological interpretation. For a comprehensive analysis of multisource data, this study first employs the Dempster–Shafer (DS) fusion technique to fuse Sentinel-2A, Landsat-9, and ASTER remote sensing images for an enriched spectral insight and improved spatial resolution. Multisource data fusion is then used to integrate the resultant fused remote sensing imagery with the digital elevation model (DEM) and Bouguer anomaly gravity data, harnessing their collective advantages for a rich diagnosis, interpretation, and identification of geological features.

The roadmap of this study, as shown in Figure 1, integrates a hybrid attention mechanism that comprises both channel attention and spatial attention with a channel separation configuration for refined channel features. The attention module minimizes the spectral and spatial redundancy of pixels by efficiently harnessing the spatial and spectral information of pixels. Consequently, the resultant attention features are fed into the LSTM network for sequential iterative processing. This results in a significant improvement in the extraction of the spectral and spatial details of target feature representations. In the end, these extracted details are fed as input into a $softmax$ function for the classification task.

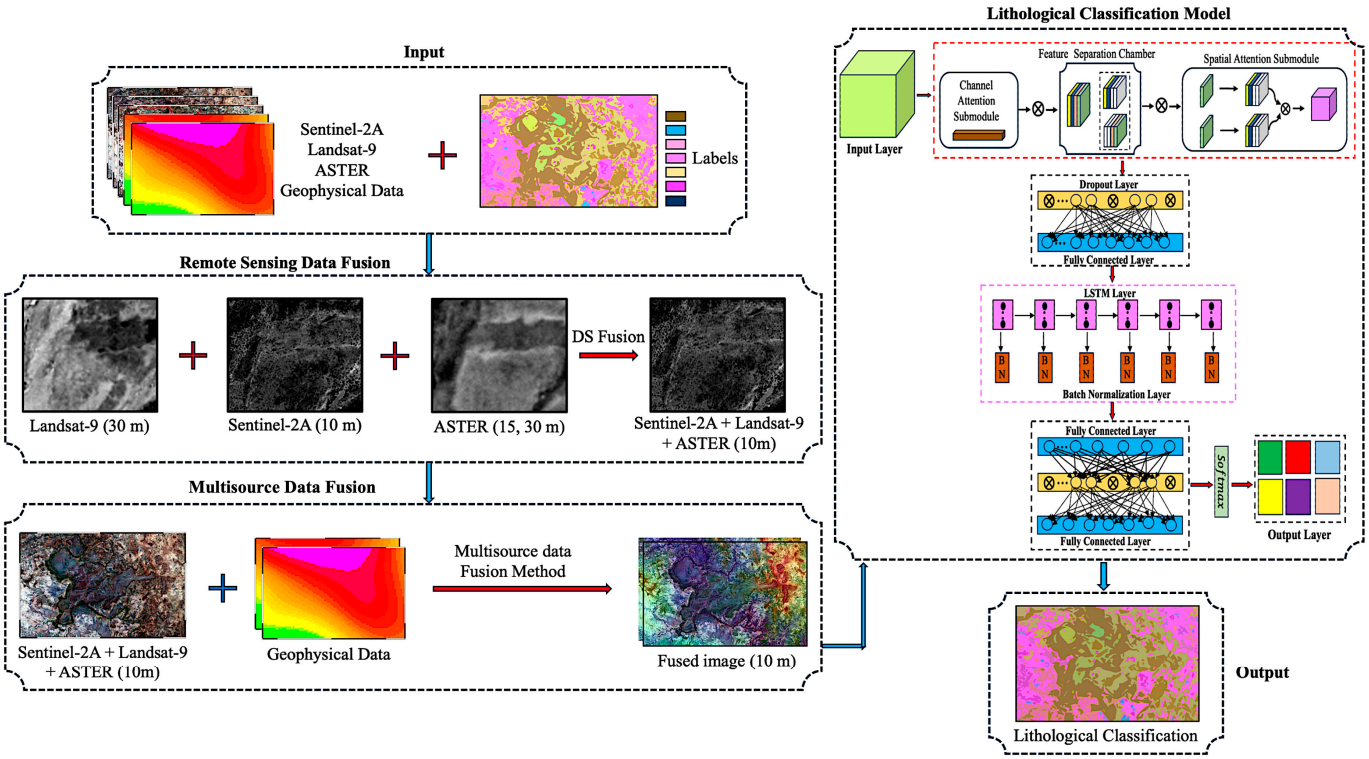

Figure 1. An overview of this study's workflow: The multisource data fusion technique is employed to fuse the gravity anomaly data and remote sensing data. Channel and spatial attention mechanisms are modeled to learn the spatial and spectral information of pixels in the fused data and the resultant attention features, fed into the LSTM network for sequential iterative processing to map lithology.

For clarity, a summary of this study's contributions is outlined as follows:

1. An improved LSTM model is proposed for efficient geological feature learning and interpretation, incorporating spectral and spatial contexts from multiple data streams.
2. A hybrid attention module with channel- and spatial-attention-based feature transformations is introduced in the LSTM network, creating a proficient HAM-LSTM. A

modeled feature separation chamber within the attention module accentuates key information and filters out less essential information. This improves information transmission in the network with reduced computational redundancy and model complexity.

3. Ablation experiments are conducted to validate the efficiency of HAM-LSTM. The model is comparatively evaluated against state-of-the-art and different attention-based models, outperforming the competing models on the multisource dataset.
4. This study corroborates the significance of integrating multiple data streams in remote sensing geological analysis and interpretation in intricate terranes, using the West African Craton as a case study.

The subsequent sections of this article are presented as follows: Section 2 presents a detailed literature review of the preliminary concepts based on which the proposed experiment for this study is designed. Section 3 describes the study area, regional background, and datasets used. Section 4 details this study's methods. The experimental design and results are outlined and analyzed in Section 5. The discussion is presented in Section 6, while Section 7 concludes this article.

2. Related Work

In this section, the preliminary concepts based on which the proposed experiment for this study is designed are presented. This includes a brief discussion of the significance of leveraging varied attention modules in a composite hybrid attention model for improved feature learning and reduced computational redundancy and model complexity.

2.1. Attention Mechanism

Integrating attention mechanisms in DL implementations is significant for model improvement. This is based on the ability of attention modules to efficiently distinguish and assign computational resources with an emphasis on prioritizing relevant information in an input dataset [22–24]. Primarily, there are four notable attention mechanisms in computer vision, namely, self-attention, channel attention, spatial attention, and hybrid attention [22,25].

Channel attention projects relevant channels in input data by generating attention masks across the channel dimension. This is epitomized by architectures such as the frequency channel attention model [26], the squeeze-and-excitation attention model [27], and the efficient channel attention model [28]. Exemplified by models such as the point-wise spatial attention network [29] and the gather–excite network [30], spatial attention models project relevant spatial domains in input data by generating attention masks across the spatial axis. Self-attention, typified by architectures such as ConvNet [31], transformer models [24,32], and visual attention models [33], generates a composite representation of a target feature in input data by expediting the relationship and contrast evaluation of input features across varied positions. Hybrid attention, however, integrates varied attention models, thereby leveraging their advantages in a composite attention architecture. For example, Li et al. [34] improved image classification by leveraging channel attention and spatial attention in a composite hybrid attention model. A composite overlapping cross-attention network has been proposed by Chen et al. [35], fusing a self-attention model and a channel attention model.

2.2. Attention Mechanism in Recurrent Neural Networks

In this study, a hybrid attention module comprising channel attention and spatial attention [34] is adapted and embedded into a recurrent neural network (RNN), precisely an LSTM, creating a hybrid-attention-based LSTM (HAM-LSTM) model. The performance of this geological interpretation model is assessed by classifying lithological features, using synthesized high-resolution multisource fusion imagery. The framework and output dimensions of the proposed hybrid-attention-based LSTM are detailed in Section 4 of this article. An overview of the general framework of attention mechanisms

in RNNs is presented in this subsection to support the understanding of this study's proposed approach.

Generally, in RNN blocks incorporated with an attention layer, the distinctive effect of varied segments in the network is assessed by the attention layer using footprints of alignment within the output feature maps. The alignment vector denoted by a_t, serving as the basis on which the impact is assessed, is computed as follows:

$$a_t(s) = align(h_t, h_s) = \frac{\exp(score(h_t, h_s))}{\sum_{s'}(score(h_t, h_s))} \tag{1}$$

In Equation (1), the current target hidden state and the entire source hidden state are denoted by h_t and h_s, respectively, at time step t. The attention mechanism within the LSTM layers is defined by the *score* function through the valuation of the corresponding correlation between h_t and h_s. Thus, in this study, the representation of the correlation is presented through the general attention as follows [36]:

$$score(h_t, h_s) = h_t^T W_s h_s \tag{2}$$

where the weight matrix is denoted by W_s, with a dismissed bias. In encapsulating information using the soft attention mechanism [37], c_t denoting the context vector is defined as follows:

$$c_t = \sum a_t(s) h_s \tag{3}$$

In the amalgamation of features from the context vectors and the input, h_t denoting the current hidden state layer is applied as follows:

$$\hat{h}_t = \tan\left(W_c \left[c_t \vdots h_t\right]\right) \tag{4}$$

With the weight of the context vectors represented by W_c, the resultant y of the attention layer is derived by feeding the concatenation layer $\hat{h}_t$ into a $softmax$ activation as follows:

$$y = softmax\left(W_t \hat{h}_t\right) \tag{5}$$

where the weight term is denoted by W_t.

2.3. The LSTM Network

The memory component in RNNs can enable the efficient learning of the nonlinearity of lithological features in remote sensing datasets given the fact that the RNN memory permits parameter sharing across the network. Within an RNN network, the mechanism entails the sequential transfer of states within the network from one moment to another through the hidden layers, thereby making RNNs suited for handling sequential characteristics with a remote sensing dataset [38,39]. Typically, the current state in one moment within the network is transferred to the state in the next moment with a transcended impact of the previous state on the current state.

RNNs are susceptible to gradient explosion and vanishing challenges from long-time-series datasets. A memory cell and three typical gates, namely, the input gate, the output gate, and the forget gate, are constituents in an LSTM internal processing unit purposed to handle the aforementioned challenges. At a current moment in the network, t, x_t, h_t, and c_t, imply the representative nodes in the hidden layer's input, output, and transitional state variables. Conversely, at time t_{t-1}, denoting a preceding moment, the representative nodes in the hidden layer's input, output, and transitional state variables are implied by x_{t-1}, h_{t-1}, and c_{t-1}, respectively. Within the network's memory cell, historical data are selectively removed by the forget gate, whereas the decision on the transfer of relevant data to the memory cell is carried out by the input gate. Historical information is however periodically stored by the memory cell in the nodes, while the output data are determined by the output gate.

In this study, sharing similarity with well-established LSTM architecture [13], the input gate, output gate, and forget gate are represented by i, o, and f, respectively. The network's weight matrix and bias are represented by w and b, respectively. In computing each gate within the network, the following Equations (6)–(11) [40] are utilized. Equation (6) expresses the computation of the forget gate f as follows:

$$f_t = \sigma\Big(w_f \cdot [h_{t-1}, x_t] + b_f\Big) \tag{6}$$

In Equation (6), the negligible data volume is determined by f_t with the weight and bias of the neurons in the forget gate f denoted by w_f and b_f, respectively, at a change in time from t_{t-1} to t. Equation (7) expresses the computation of the input gate i as follows:

$$i_t = \sigma_l(w_i \cdot [h_{t-1}, x_t] + b_i) \tag{7}$$

In Equation (7), the input data volume is determined by i_t, with the weight and bias of the neurons in the input gate i denoted by w_i, and b_i, respectively, at a change in time from t_{t-1} to t. Hence, the updated intermediate state variable c_t is expressed in Equation (8) at the current time as follows:

$$c_t = f_t \cdot c_{t-1} + i_t \cdot \widetilde{c}_t \tag{8}$$

$$c_t = \tan h(w_c \cdot [h_{t-1}, x_t] + b_c) \tag{9}$$

$\widetilde{c}_t$ denotes the input process of neurons at time t, whereas the weight and bias of the hidden layer are denoted by w_c, and b_c, respectively, at a change in time from t_{t-1} to t. Equation (10) expresses the computation of the output gate o as follows:

$$o_t = \sigma(w_o \cdot [h_{t-1}, x_t] + b_o) \tag{10}$$

Sequentially, the current neural output is therefore defined as follows:

$$h_t = o_t \cdot \tan h(c_t) \tag{11}$$

In Equation (11), the current output's dimension h_t is determined by the output gate o_t, through the memory unit c_t. Within the hidden layer, the neurons of the forget gate's weight and bias are denoted by w_o and b_o, respectively, at a change in time from t_{t-1} to t. In a definite equation of $\sigma(*) = 1/(1 - e^{-x})$, $\sigma(*)$ implies a sigmoid function, whereas in $\tan h(*) = (e^x - e^{-x})/(e^x + e^{-x})$, $\tan h(*)$ represents the $\tan h$ function.

From the defined attributes, it is quite palpable that the LSTM network is capable of handling continuous and significantly correlated time-series data. Nonetheless, in mapping lithology, learning lithological features at a current moment is highly dependent on the underlying characteristic features of the sample dataset. In implementing the LSTM network in the experiments, a basis for the prediction and onward classification of lithological units in the multisource imagery is created by establishing a full-bodied correlation with a competent means of feedback from the sample data.

3. Study Area and Data

3.1. Regional Geological Setting

The study area is located in southern Burkina Faso in West Africa stretching between 10.65° to 11.1°N and 2.75° to 3.2°W. It spans an area of 729 km^2 across various towns and villages including Dissin, Bélarkaré, Diebougou, and Navielgane. The land surface of the study area is predominantly semi-arid with sparse vegetation. The main geological units of the study area are the Paleoproterozoic and associated greenstone belts (Figure 2). It is part of the West African Craton, notably amongst the oldest geological formations on planet Earth [41]. It has an intricate and extensive geological background stretching over 3 Ga [42].

As a massive geological unit, the craton spans an area of about 3 million square kilometers. Several distinctive geological regions constitute this massive geological unit. These geological regions include the Kénéma-Man Shield and Reguibat Shield in the south and north, respectively [43]. This craton's existence is a result of three key tectonic events in geological history: the Pre-Leonean Orogeny (3.5–2.9 Ga), responsible for the craton's oldest outcrops; the Liberian Orogeny (2.9–2.8 Ga), emanating from the convergence between an early island arc system and the Archean craton; and the Pre-Eburnean (~2.19–2.15 Ga) and the Eburnean Orogeny (2.15–1.8 Ga), which characterized the formulation of associated metallogenic potentials within the volcano-sedimentary rocks, volcanic rocks, and sedimentary rocks of the Birimian Supergroup [41].

Different lithological units and geological features within the craton contribute to the formation of mineral resources in the craton. The occurrence of intrusive formations during the Cretaceous era influenced distinctive mineralization in the geological province significantly. Within the West African Craton, the Birimian Supergroup is noted for substantial gold mineralization with a well-documented mining history, chiefly in the Birimian Greenstone belt and the Ashanti belt.

Beyond mineralization in the Birimian Supergroup, outstanding deposits of minerals including precious diamond stones, nickel, copper, iron ore, manganese, uranium, lead, zinc, aluminum ore, cobalt, platinum, and phosphate are hosted in the craton by the Kénéma-Man Shield in the Man domain, the Baoulé-Mossi domain, and the Kédougou-Kéniéba Inlier (KKI) [43].

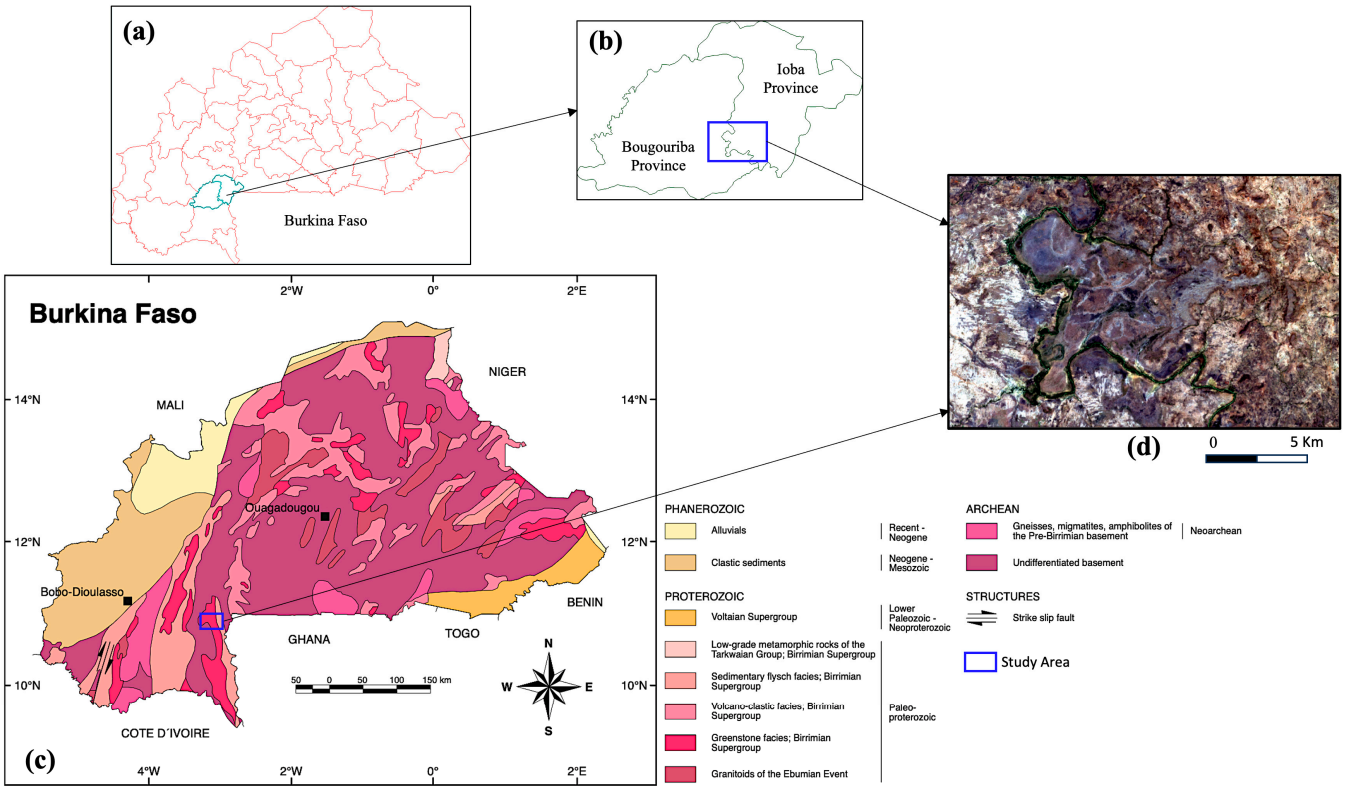

Figure 2. Location of study area and regional geological setting. (**a**) Administrative map of Burkina Faso; (**b**) administrative map of Bougouriba and Ioba Provinces within which the study area is located; (**c**) geological overview of Burkina Faso (modified from [44]) indicating the study area; (**d**) color composite image of Landsat-9 covering the study area.

3.2. Multisource Data

This study integrates data from five data sources. It leverages the multispectral bands of Landsat-9, Sentinel-2A, ASTER's notable spectral resolution [45], Advanced Land Ob-

servation Satellite (ALOS) Phased Array type L-band Synthetic Aperture Radar (PALSAR) DEM, and 100 m (m) gravity-sampling-spaced EGM2008 Bouguer anomaly data. This combination facilitates efficient lithological mapping by offering detailed spectral insight across the visible near-infrared (VNIR), shortwave infrared (SWIR), and TIR spectrum of the electromagnetic wavelength.

The ALOS PALSAR DEM data provide a detailed topographic insight into the geological features in the study area. The Bouguer anomaly data provide details about the study area's subsurficial geological structures, regional geological context, lithological variations, and depth estimation. The two polar-orbiting Sentinel-2A satellites capture spectral data in 13 bands. Spectral insights from bands 2, 3, and 4 and the near-infrared band 8 of Sentinel-2A's secondary subsurface reflectance data with 10 m resolution are used in this study to synthesize fusion imagery with Landsat-9 and ASTER imagery. The spectral resolution of the ASTER data used spans nine VNIR and SWIR bands at 15 m and 30 m, respectively.

Landsat-9 [46] with its Operational Land Imager-2 (OLI-2) captures data in nine spectral bands at a 30 m resolution excluding its 15 m panchromatic band. Landsat-9's thermal infrared sensor-2 (TIRS-2) provides two spectral bands at 100 m. Boasting a 14-bit radiometric resolution, Landsat-9 can distinguish between 16,384 shades of any particular wavelength, outclassing Landsat-8's 4096 at 12 bit. A clearer identification of patterns in imagery is easier in Landsat-9 aided by its improved resolution. This augments the process of labeling training samples given the enhanced brightness contrast of distinctive features.

The Landsat-9, ASTER, and ALOS PALSAR DEM data used were generated in Level 1T and captured on 20 January 2022, 24 April 2015, and 28 August 2007, respectively. The products were then delivered by the United States Geological Survey (https://earthexplorer.usgs.gov/, accessed on 28 October 2022). The Sentinel-2A data were captured on 20 March 2023 and provided on the European Space Agency's Copernicus Data Space Ecosystem (https://dataspace.copernicus.eu/, accessed on 28 October 2022). The gravity data were acquired from the GFZ—German Research Centre for Geosciences (www.gfz-potsdam.de, accessed on 28 October 2022).

3.3. Hardware and Software

The experimental set in this study was conducted using an Apple MacBook Pro laptop computer (Apple Inc, Cupertino, CA, USA). The laptop computer has an Apple silicon M1 chip with a 32-core GPU, 10-core CPU, 16-core Neural Engine, 32 GB RAM, and a memory bandwidth of 400 GB/s.

In preprocessing the remote sensing images, ENVI 5.3 and QGIS 3.34 software were utilized. Geosoft Oasis Montaj 2023.1 software was utilized in preprocessing the gravity anomaly data. Subsequently, Python 3.12.4. was used to perform data processing, model training, and testing, evaluation, and postprocessing tasks. In performing various tasks, Python libraries and modules including rasterio, itertools, geopandas, torch, matplotlib, numpy, PyTorch, sklearn, pycm, pickle, and others were employed.

3.4. Data Preprocessing

Using ENVI, radiometric calibration was applied on the Landsat-9 product to deduce the most prominent reflectance, temperature brightness, or radiance for the image obtained. Subsequently, First-Line-of-Sight Atmospheric Analysis of Spectral Hypercube (FLAASH) was employed to rectify atmospheric distortion in the image. The FLAASH algorithm as an image correction tool corrects wavelengths up to 3 μm within the VNIR and SWIR ranges of the electromagnetic spectrum based on a first-principles atmospheric correction code.

Atmospheric correction was performed on the Sentinel-2A product and the VNIR and SWIR bands, stacked to generate a 10 m composite band dataset for processing. In preprocessing the ASTER product, atmospheric transformation, radiation correction, and orthophoto transformation were applied. The SWIR bands of the ASTER product were resampled to equate the 15 m resolution of the VNIR bands. The effect of radiation

errors on image distortion was addressed by applying FLAASH to eliminate the effects of atmospheric scattering and absorption in the scene.

In tackling the effect of topography on the ASTER sensor's information capture, orthophoto transformation using ENVI's RPC orthorectification module was performed to conduct topographic relief operations in the scene. Performing these preprocessing tasks enhances target feature identification in hilly environments within the scene. Using Geosoft Oasis Montaj, topographic correction and Butterworth filtering were performed on the gravity anomaly data to generate a complete Bouguer anomaly (CBA) map. After the CBA map, Gaussian regional and residual filtering was performed on the enhanced CBA map to generate a residual gravity map using Geosoft Oasis Montaj software [47,48].

The vegetation cover in the Landsat-9, Sentinel-2A, and ASTER scenes was subsequently masked. The initial input dataset for integrating the multiple remote sensing datasets spanned twenty (20) bands: four (4) bands from Sentinel-2A, seven (7) bands from Landsat-9, and nine (9) bands from ASTER. Following the implementation of the Dempster–Shafer (DS) fusion, bands that provided complimentary spectral and spatial information were retained to reduce redundancy. This resulted in the retention of seven spectral bands with 10 m resolution. The multisource data fusion technique was subsequently employed to integrate these seven bands with each element of the Bouguer anomaly data (100 m) and ALOS PALSAR DEM data (12.5 m), based on respective correlations between the spectral reflectance and geophysical element concentration or topographic intensity. This fusion resulted in a final composite image with three bands that captured the most informative spectral and geophysical characteristics. The final fused image increases the spatial resolution of both the Bouguer anomaly data and the ALOS PALSAR DEM data from 100 m and 12.5 m to 10 m, respectively, providing additional diagnostic insight for lithological mapping, as shown in Figure 3.

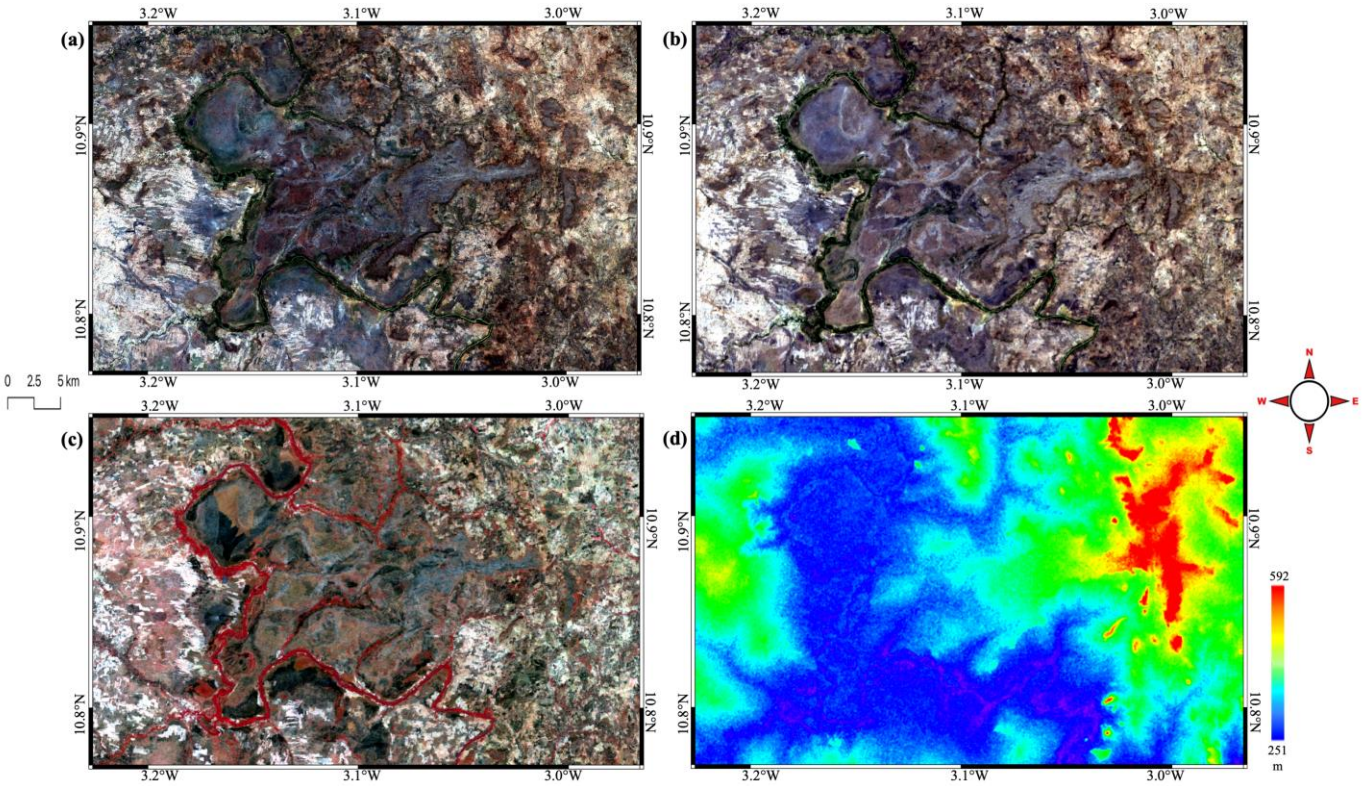

Figure 3. False color composite imagery of remote sensing data used: (**a**) Sentinel-2A (bands 4-3-2); (**b**) Landsat-9 (bands 4-3-2); (**c**) ASTER (bands 3-2-1); and (**d**) 12.5 m spatial resolution high-precision ALOS PALSAR DEM.

4. Methodology

4.1. Methodology for Integrating Multisource Data for Lithological Classification

4.1.1. Masking Vegetation in the Remote Sensing Scenes

In processing the satellite imagery, vegetation masks of the semi-arid Sentinel-2A, Landsat-9, and ASTER scenes were generated by contrasting pixel reflectance, filtering zones with appreciable vegetation cover. As illustrated in Figure 4, utilizing the normalized difference vegetation index (NDVI) [49], vegetation coverage from the NDVI values was evaluated ranging from 5% to 95% confidence, as presented in Equations (12) and (13):

$$NDVI = \frac{R_{NIR} - R_R}{R_{NIR} + R_R} \tag{12}$$

$$f_{veg} = \frac{NDVI - NDVI_{Min}}{NDVI_{Max} + NDVI_{Min}} \tag{13}$$

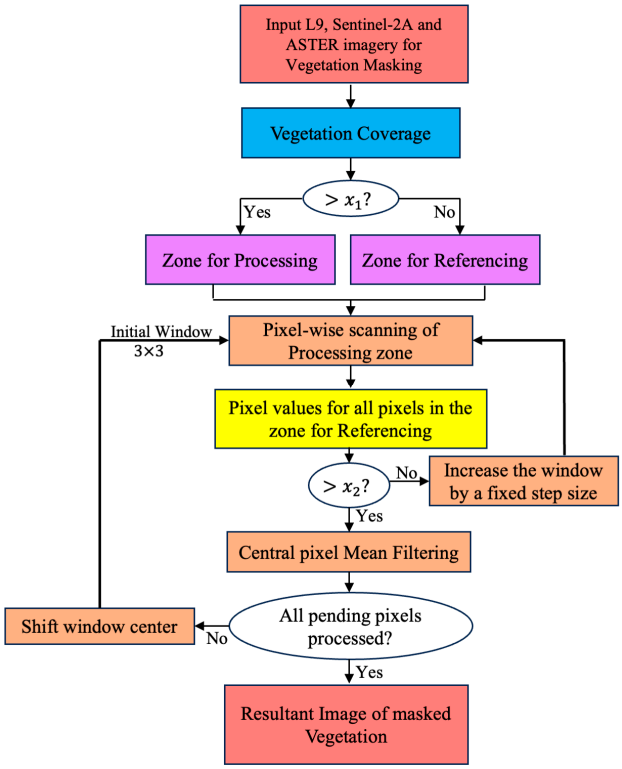

Figure 4. Vegetation masking workflow.

The reflectance values of the remote sensing imagery's near-infrared band (band 5 for Landsat-9; band 8 for Sentinel-2A; and band 3 for ASTER) and the red band (Landsat-9's band 4; band 4 for Sentinel-2A; and ASTER's band 2) are denoted by R_{NIR} and R_R, respectively. The 5th percentile value of the cumulative NDVI distribution is denoted by $NDVI_{Min}$, whereas $NDVI_{Max}$ denotes the 95th percentile value. The vegetation cover is denoted by f_{veg}. To determine the cumulative NDVI distribution, all pixel NDVI values were ranked, followed by a pixel-wise computation deducing the cumulative NDVI distribution histogram.

Subsequently, a definite vegetation coverage threshold, x_1, of 65% was employed to divide the scene into zones characterized by low and high vegetation cover. The 65% threshold was selected as a mid-range filter to distinguish the areas with moderately dense vegetation in the scene from areas with denser vegetation cover. A similar approach was used by J. Lu et al. [50]. With this approach, areas where the vegetation cover was not dense enough to distort the spectral reflectance of the underlying geological features were factored

into the analysis. Given the semi-arid conditions of the scene, the substantial vegetation cover in the area was effectively distinguished and masked by applying this threshold without over-masking regions with sparse vegetation, thus creating a balance in preserving detail and providing clarity for onward geological feature analysis and interpretation.

These zones characterized by low and high vegetation cover were classified as the "processing zone" and "referencing zone", respectively. In the end, a pixel-wise mean filtering operation was conducted on the processing zone. This involved the substitution of target pixel values within the processing zone with the mean value of neighboring pixels from the reference zone within a designated filtering window. This was governed by the condition that within the filtering window the number of reference pixels does not fall below the threshold x_2 of value 10.

Masking the vegetation in the scene was not only essential for improving data quality but also enhanced consistency in image analysis across all the different satellite imagery used in this experiment. Using this approach improves geological feature detection and aids in accurate feature interpretation in the dataset.

4.1.2. Dempster–Shafer (DS) Fusion Technique

In fusing the Sentinel-2A, Landsat-9, and ASTER images, the Dempster–Shafer (DS) fusion technique [51,52] was employed in this study. Guided by multiple evidence theory, the DS fusion method evaluates the probability of hypotheses by employing probability and uncertainty intervals [53,54]. The possibility of each hypothesis being true is computed, with the probability and uncertainty metrics producing the same scores where hypotheses are mutually congruent.

The DS transform was used to synthesize a high-resolution color image from the spectral details of the Sentinel-2A, Landsat-9, and ASTER images. The majority of the spectral details of the input images were retained in the resultant fused imagery with improved image quality, thus providing clearer contrast for the identification of ground features. The basic principles of the DS fusion method are presented as follows:

Given a set of image classes denoted by a frame of discernment θ, the primary mass function is deduced as follows:

$$\mathrm{x} : 2^{\theta} \rightarrow [0,1]; \sum_{\mathrm{F} \subset 2^{\theta}} \mathrm{x}(\mathrm{F}) = 1; \mathrm{x}(\phi) = 0 \tag{14}$$

The volume of substantiation is set by the belief function denoted by $\mathrm{Bel}(\mathrm{F})$, which infers the interpretation of F. Based on the frame of discernment, the function is expressed as follows:

$$\mathrm{Bel}(\mathrm{F}) = \sum_{\mathrm{C} \subset \mathrm{F}} \mathrm{x}(\mathrm{C}) \tag{15}$$

The function deducing the plausibility of the operation is denoted by $\mathrm{Pl}(\mathrm{F})$, which implies the volume of substantiation and thus supports F, as follows:

$$\mathrm{Pl}(\mathrm{F}) = \sum_{\mathrm{C} \cap \mathrm{F} \neq 0} \mathrm{x}(\mathrm{C}) \tag{16}$$

The plausibility function $\mathrm{Pl}(\mathrm{F})$ can thereby be inferred from the belief function $\mathrm{Bel}(\mathrm{F})$ as follows:

$$\mathrm{Pl}(\mathrm{F}) = 1 - \mathrm{Bel}(\mathrm{F}) \tag{17}$$

The DS theory integrates evidence from varied sources. The integration operation, known as the orthogonal sum, is expressed as follows:

Given $\mathrm{K} \neq 1$, and $\forall \mathrm{A} \subseteq \theta$, $\mathrm{A} \neq \phi$,

$$(\mathrm{x}_1 \oplus \mathrm{x}_2)(\mathrm{A}) = \frac{\sum_{\mathrm{F}_1 \cap \mathrm{F}_2 = (\mathrm{A})} \mathrm{x}_1(\mathrm{F}_1)\mathrm{x}_2(\mathrm{F}_2)}{1 - \mathrm{K}} \tag{18}$$

$$K = \sum_{F_1 \cap F_2 = (\phi)} x_1(F) x_2(F_2) \tag{19}$$

$$(x_1 \oplus x_2)(\phi) = 0 \tag{20}$$

Here, $\oplus$ denotes element-wise summation. The $\oplus$ operator adds information from the varied input images, thereby generating the refined fusion image. The conceptual implementation of the Dempster–Shafer fusion technique to fuse the Sentinel-2A, Landsat-9, and ASTER images is illustrated in Figure 1.

4.1.3. Multisource Data Fusion

The resultant fusion image from the remote sensing images was fused with the gravity anomaly geophysical data using a multisource data fusion technique. The multispectral bands of the remote sensing imagery and the layers of the geophysical unit cover the study area and align within the same spatial coordinate system. In fusing the geophysical data and remote sensing imagery, the multisource data fusion workflow outlined by Ding et al. [55] was adopted in this study as follows:

1. The multispectral remote sensing fusion imagery (MS) and the single-layer geophysical data (geo_l) were clipped to an equivalent size. For a guaranteed alignment, spatial correction and registration were then performed on the input datasets.
2. The spatial and background information of the input multispectral image was deduced, which was denoted by the high-frequency component (MS_H_i) and low-frequency component (MS_L_i) of the multispectral image, respectively. These frequency components were generated by decomposing the multispectral image, using the Laplacian pyramid algorithm [56,57].
3. The cubic convolution technique was employed to resample the low-frequency component (MS_L_i) of the multispectral bands and the layered geophysical data $\left(geo_l_j\right)$ to match the resolution of the multispectral bands' high-frequency component (MS_H_i) [58].
4. A relationship was defined between the geophysical measurements and the spectral reflectance. The establishment of this relationship entailed the development of a relational function to deduce the correlation coefficient (f_{ij}), which is defined as follows:
$$geo_l_j = f_{ij} \cdot MS_L_i \tag{21}$$
where the band index and geophysical parameter index are represented by i and j, respectively.
5. The correlation coefficient (f_{ij}) was then used to generate a high-frequency information layer (geo_h), resulting from the reconstruction of the geophysical layer and the multispectral bands' high-frequency component (MS_H_i).
$$geo_h = f_{ij} \cdot MS_{Hi} \tag{22}$$
6. A fused high-resolution geophysical layer (geo_f) was produced from the injection of the resultant high-frequency information layer, representing spatial details, into the resampled geophysical layer.
$$geo_f = geo_{lj} + geo_h \tag{23}$$

4.2. Hybrid Attention Module

Figure 5 depicts the hybrid attention module. Given a transitional feature map input of $I \in \mathbb{R}^{H \times W \times C}$, a one-dimensional channel attention map $A_c \in \mathbb{R}^{1 \times 1 \times C}$ is inferred by HAM. This one-dimensional channel map is then multiplied with the input feature map, deriving a refined channel-defined feature. Subsequently, a channel separation operation is executed by the feature separation chamber to allot refined channel features into two

clusters based on channel attention maps by the channel dimension. The resultant two clusters of refined channel features are then fed into the spatial attention submodule. As illustrated in Figure 5, 2D spatial attention maps are thereby generated by the following two groups: $A_{s,1} \in \mathbb{R}^{H \times W \times 1}$ and $A_{s,2} \in \mathbb{R}^{H \times W \times 1}$.

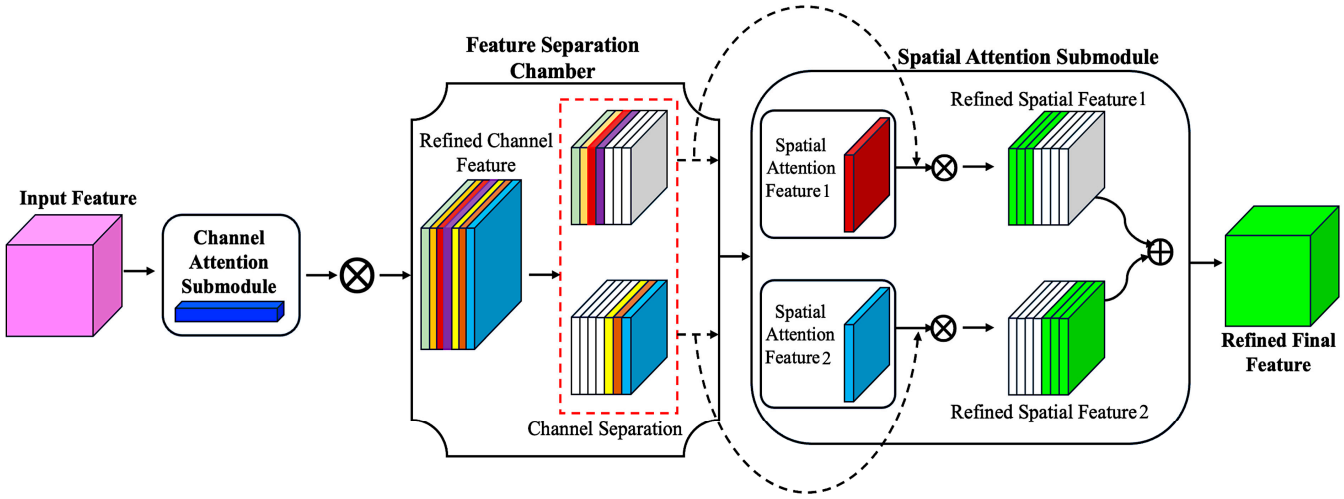

Figure 5. The HAM structure. It comprises three sequential components: channel attention submodule, feature separation chamber, and spatial attention submodule. One-dimensional and two-dimensional feature maps are produced by the channel and spatial attention submodules, respectively.

Thereafter, a pair of refined spatial features is produced by the spatial attention submodule via the multiplication of generated maps with their collective features. By concatenating the collective pair of spatial features, the refined final feature is generated. The following equations depict this process:

$$I' = A_c(I) \otimes I \tag{24}$$

$$I'_1 \oplus I'_2 = I' \tag{25}$$

$$I''_1 = A_{s,1}(I_1) \otimes I'_1 \tag{26}$$

$$I''_2 = A_{s,2}(I'_2) \otimes I'_2 \tag{27}$$

$$I'' = I''_1 \oplus I''_2 \tag{28}$$

Here, $\otimes$ denotes element-wise multiplication, and $\oplus$ denotes element-wise summation. The $\otimes$ operator is used to calculate the channel attention values and spatial attention values. The refined channel feature I' is generated by the broadcast of the resultant channel attention values along the channel dimension. Likewise, the refined spatial feature I'' is generated by the channel-wise broadcast of the resultant values of spatial attention. The $\oplus$ operator, however, adds the two refined features, thereby generating the final feature output I''. This is made possible by splitting the refined channel feature I' into two separate groups, I'_1 and I'_2, along the channel dimension. Each attention submodule of the HAM module is detailed in subsequent sections, with graphical illustrations of their respective attention map computations in Figures 6 and 7.

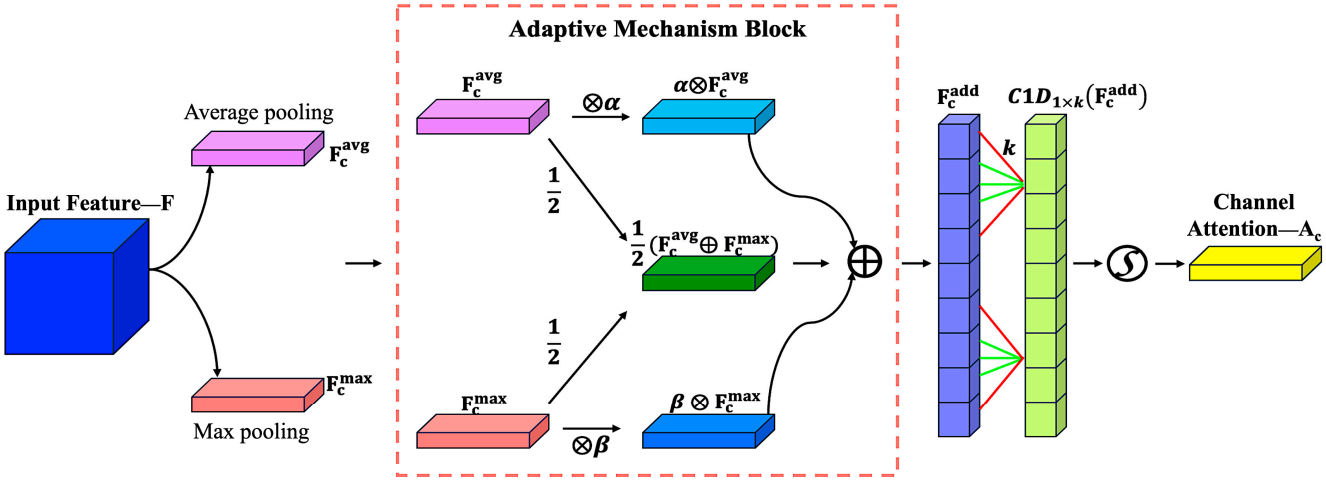

Figure 6. Framework of HAM's channel attention submodule. Dimensional feature information is generated by both max-pooling and average-pooling operations. The resultant features are then fed through a one-dimensional convolution with a sigmoid activation to deduce the definitive channel feature.

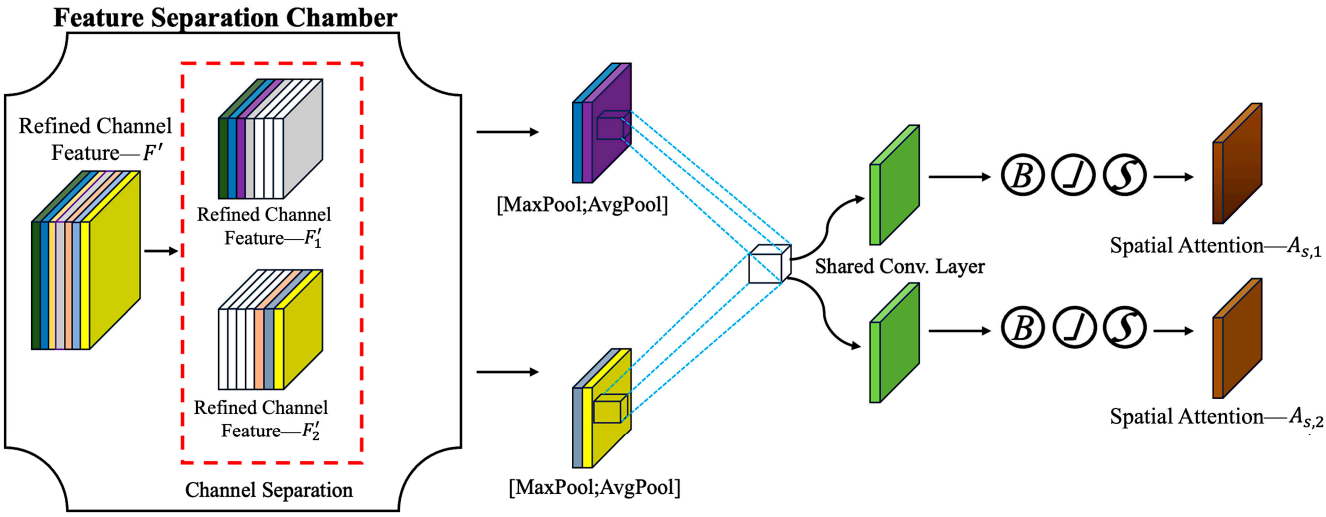

Figure 7. Framework of HAM's spatial attention. Two feature clusters of partitioned refined channel features from the separation chamber are fed into the submodule. Average-pooling and max-pooling functions subsequently synthesize two pairs of 2D maps into a shared convolution layer to synthesize spatial attention maps.

4.2.1. Channel Attention Submodule

As Figure 6 depicts, spatial dimension information is accumulated in the channel attention by employing both max-pooling and average-pooling operations. This results in the generation of two feature maps, I_c^{avg} and I_c^{max}. These generated feature maps are then referred to the adaptive mechanism block for onward feature enrichment, $I_c^{add} \in \mathbb{R}^{1\times1\times C}$. Two parameters, α and β, are present in the adaptive mechanism block and are trainable using stochastic gradient descent (SGD). The computations in the adaptive mechanism block are summarized in Equation (29).

$$I_c^{add} = \frac{1}{2} \otimes \left(I_c^{avg} \oplus I_c^{max} \right) \oplus \alpha \otimes I_c^{avg} \oplus \beta \otimes I_c^{max} \tag{29}$$

where α and β, as components of the average-pooled and max-pooled features, are two floating variables greater than zero and less than one in the adaptive mechanism. Beyond feature enrichment, these parameters also improve the image feature extraction process.

Using a fast one-dimensional convolution, interactions between channels are captured with a set kernel size k. This represents the interactions between k neighbors. The value of k can be determined by Equation (30) as follows:

$$k = \phi(\mathrm{C}) = \left| \frac{\log_2(\mathrm{C})}{\gamma} + \frac{b}{\gamma} \right|_{odd} \tag{30}$$

where $|v|_{odd}$ implies the closest odd numeral t, with C representing the channel sum; the adaptive kernel size k can be deduced by implementing a mapping function ϕ. In this study, the hyperparameters b and γ are set to 1 and 2, respectively. The resultant feature map $\mathrm{I_c^{add}} \in \mathbb{R}^{1\times1\times\mathrm{C}}$ is subsequently fed through a one-dimensional convolution with a sigmoid activation function. Equation (31) gives a general computation for the channel attention implementation.

$$\begin{aligned} \mathrm{A_c(I)} &= \sigma\Big(C1D_{1\times k}\Big(\tfrac{1}{2} \otimes (\mathrm{AvgPool\ (I)} \oplus (\mathrm{MaxPool\ (I)} \oplus \alpha \otimes \\ &\quad (\mathrm{AvgPool\ (I)} \oplus \beta \otimes (\mathrm{MaxPool\ (I)}\Big)\Big) \\ &= \sigma\Big(\mathrm{C1D_{1\times k}}\Big(\tfrac{1}{2} \otimes \Big(\mathrm{I_c^{avg}} \oplus \mathrm{I_c^{max}}\Big) \oplus \alpha \otimes \mathrm{I_c^{avg}} \oplus \beta \otimes \mathrm{I_c^{max}}\Big)\Big) \\ &= \sigma\Big(\mathrm{C1D_{1\times k}}\Big(\mathrm{I_c^{add}}\Big)\Big) \end{aligned} \tag{31}$$

where σ denotes the sigmoid activation function, whereas $\mathrm{C1D_{1\times k}}$ denotes a one-dimensional convolution with k kernel size.

4.2.2. Channel-Feature Separation

Unique weights are assigned to distinctive channels for the onward partitioning of refined features into two channel clusters along the channel dimension. This partitioning process is termed channel separation.

The channel-feature separation process is based on a separation rate hyperparameter λ, which distinguishes between the most relevant and less relevant channel clusters. The channel dimension of relevant features is determined by computing the refined channel feature with λ. The closest relevant feature, denoted by C_{im}, is selected based on the channel dimension's even numerical property. Equation (32) expresses the C_{im} computation as follows:

$$C_{im} = |C_{C-R} \cdot \lambda|_{even} \tag{32}$$

In Equation (32), the dimension of the refined channel features is denoted by C_{C-R}, whereas $|v|_{even}$ denotes v's nearest even number. C_{im} denotes important channel features. The values of the channel dimension of the refined important and sub-important channel features are defined as even numbers, which are ensured by C_{im}. Within the channel attention map, the top n maximum values, corresponding to the top n important channels, are identified by C_{im} in the refined feature set. However, it is worth noting that C_{im} and n differ in their physical interpretation despite their numerical equivalence. In the computation stream, the largest n^{th} value in the channel attention map can be deduced once the determinant n value has been identified. Subsequently, the values within the attention map can then be categorized into two as follows: values equivalent to or greater than the nth largest value, classified as the important part, and values lower than the nth largest value, classified as the sub-important part.

This distinction is operationalized by defining two masks, sharing the same shape as the channel attention map. A value of 1 is assigned to the important part, and a 0 value is assigned to the sub-important part by the first mask, whereas the second mask assigns a value of 0 to the important part and a value of 1 to the sub-important part. Applying these masks termed important mask and sub-important mask in an element-wise multiplication to the refined channel feature effectively segments the refined channel feature into two

groups, namely, the important channel group F_1' and the sub-important channel group F_2'. As highlighted in Figure 7, the white cuboids represent the zero values of the refined channel features F_1' and F_2'. Within the channel separation chamber, the channel separation operation retains the channel dimension of the refined channel features and thus satisfies Equation (25).

4.2.3. Spatial Attention Submodule

Two sets of feature maps $A_{s,1} \in \mathbb{R}^{H\times W\times 1}$ and $A_{s,2} \in \mathbb{R}^{H\times W\times 1}$ are subsequently generated by integrating a 7×7 convolution layer, denoted by $C2D_{7\times 7}$ on the feature descriptors. These indicate where to suppress or emphasize.

The processes within the submodule are clarified further in Figure 7. The two channel clusters from the feature separation chamber, I'_1 and I'_2, representing the important feature group and less important feature group, respectively, from the refined feature I', are fed into the submodule. Average-pooling and max-pooling functions subsequently synthesize two pairs of 2D maps from I'_1 and I'_2. One set is denoted by $I_{s,1}^{avg} \in \mathbb{R}^{H\times W\times 1}$ and $I_{s,1}^{max} \in \mathbb{R}^{H\times W\times 1}$, whereas the other set is represented by $I_{s,2}^{avg} \in \mathbb{R}^{H\times W\times 1}$ and $I_{s,2}^{max} \in \mathbb{R}^{H\times W\times 1}$.

The computation of the spatial attention process can be generalized by Equations (33) and (34).

$$\begin{aligned} A_{s,1}(I') &= \Phi\left(C2D_{7x7}\left(\left[AvgPool(I'_1); MaxPool(I'_1)\right]\right)\right) \\ &= \Phi\left(C2D_{7x7}\left(\left[I_{s,1}^{avg}; I_{s,1}^{max}\right]\right)\right) \end{aligned} \tag{33}$$

$$\begin{aligned} A_{s,2}(I') &= \Phi\left(C2D_{7x7}\left(\left[AvgPool(I'_2); MaxPool(I'_2)\right]\right)\right) \\ &= \Phi\left(C2D_{7x7}\left(\left[I_{s,2}^{avg}; I_{s,2}^{max}\right]\right)\right) \end{aligned} \tag{34}$$

In the aforementioned Equations, Φ implies sequential nonlinear conversions. These conversions comprise sigmoid and ReLU functions and batch normalization. To remove negative values from the spatial attention feature maps, the ReLU function is used, stressing features whose impact positively influences the final classification output.

4.3. Proposed Hybrid Attention–LSTM Model Framework

The structural framework of the proposed model is presented in Figure 8. HAM-LSTM integrates a hybrid attention mechanism (HAM) which comprises both channel attention and spatial attention with a channel separation configuration for refined channel features. The attention module minimizes the spectral and spatial redundancy of pixels by efficiently harnessing the spatial and spectral information of pixels. Consequently, the resultant attention features are fed into the LSTM for sequential iterative processing. This results in the extraction of spectral and spatial details of target feature representations. In the end, these extracted details are fed as input into a $softmax$ function for the classification task.

The main procedures in this model are as follows:

1. A two-dimensional matrix is generated as model input from the augmentation, fusion, and reshaping of signals from the Sentinel-2A, Landsat-9, and ASTER sensors; ALOS PALSAR DEM, and Bouguer anomaly gravity data.
2. The LSTM cells serving as a vital component within the LSTM network ensure maximum information transmission.
3. HAM, comprising both channel and spatial attention, is embedded into each LSTM cell. In the proposed model, the attention mechanisms are implemented during feature learning in the concatenation of feature maps into the dropout layers.
4. As part of the transition layer, the attention layers can highlight sensitive features and retain distinctive features to pass onward into subsequent layers in the network.
5. In the end, a $softmax$ function is used for the classification of lithological units.

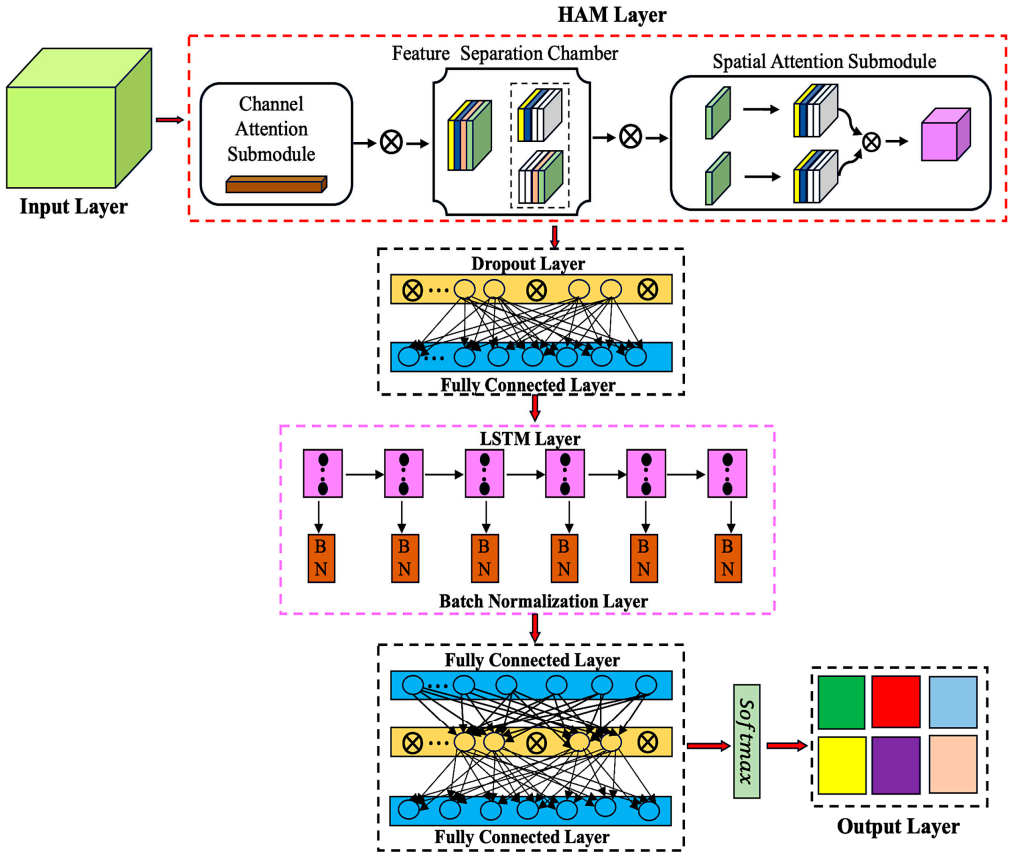

Figure 8. The structural framework of the proposed HAM-LSTM model.

4.4. Model Parameter Tuning and Output Dimensions

In the model design (Figure 8), a batch normalization layer is proposed to mitigate the potential change in input data distribution within the network; thus, it is employed in succession to the LSTM layer in the feed-forward implementation. In the model framework, the HAM module enhances the extraction of in-depth feature information for an efficient output in the model's classification performance. The dropout layer is tasked with handling overfitting in the model. By learning lithological features at a current moment based on the underlying characteristic features of the sample dataset at the preceding moment, the LSTM model establishes a basis for the prediction and onward classification of lithological units in the fused multisource data.

In training HAM-LSTM, the loss function L_f was reduced via parameter optimization, thereby augmenting the mean squared error selection for delineating lithological units. The loss function L_f, given predicted values of Y' and the equivalent real values of Y, in the experimental set is expressed as follows:

$$L_f = \frac{1}{n}\sum(Y' - Y)^2 \tag{35}$$

n in the above equation denotes the number of labels. Backpropagating the gradient in the experimental set ensures the optimization of the parameters in the HAM-LSTM model. To accelerate the training of the model, the adaptive moment estimation (Adam) [59] is utilized as the optimization function, which adjusts the learning rate l_r, dependent on the historical gradient information.

In the experimental set, the model was trained for 100 epochs. The learning rate l_r was set at 10^{-3}, with a cosine decay function of 1×10^{-6} over 10 epochs until the convergence of the model's loss function. A default *epsilon* value of 1×10^{-8} was added to curb potential numerical instability and improve the optimization process. The test dataset was

then utilized subsequently to quantitatively assess the model's efficiency in identifying lithological units.

5. Experiment and Result Analysis

5.1. Gravity Anomaly Results

The resultant CBA and residual gravity maps of the terrane are projected in Figure 9. Despite the subsurficial insight provided, the anomalous signatures depicted in the CBA map are induced by deep regional gravity anomalous effects. In contrast, these deep regional gravity anomalous effects are minimal in the residual gravity map with the projection of local geological features with better clarity. The gravity signatures of varied lithological formations in the terrane are delineated, including Birimian metavolcanics characterizing high gravity signatures ranging from 1.1 mGal to 2.1 mGal, characterized by their maficity and iron enrichment. Also, the characteristic low to moderate gravity response of metasediments ranging from 0.2 mGal to 1.1 mGal is delineated.

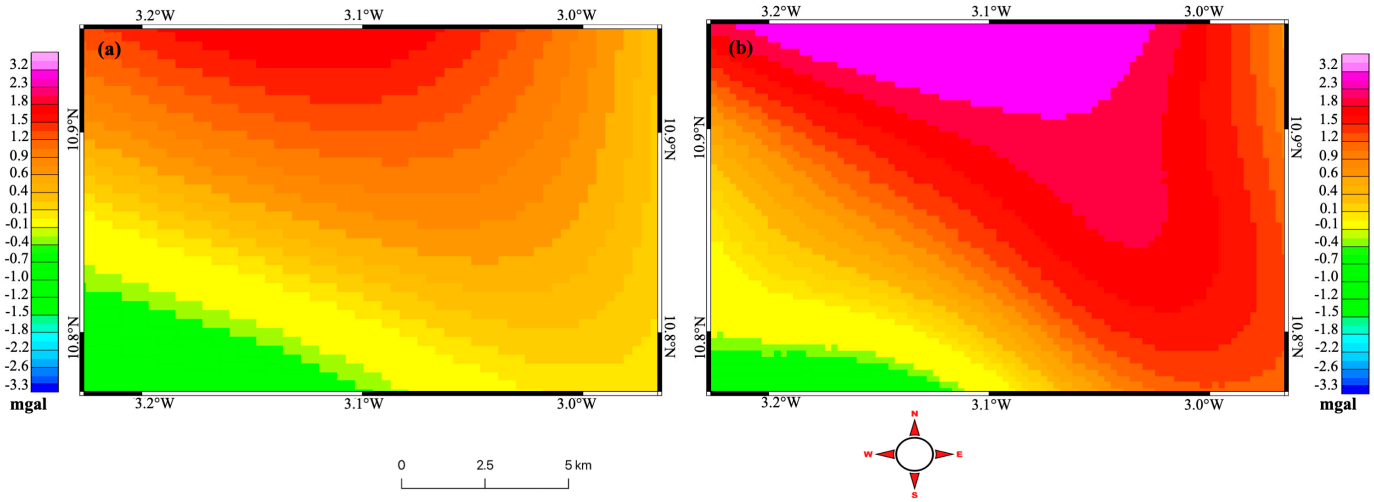

Figure 9. Gravity anomaly maps of the terrane used: (**a**) complete Bouguer anomaly; (**b**) residual gravity.

The gravity anomaly maps not only distinguish the boundaries between the varied lithological formations in the scene but also offer appreciable insight that augments the interpretation of near-surface features in the scene.

5.2. Fusing Multisource Data

For an efficient diagnosis of feature representations of target lithologies in the area of interest, employing multisource data fusion harnesses the capabilities of the varied datasets sourced for this experiment. Utilizing this fusion technique heightens information capture in the multisource datasets and improves feature diagnosis and interpretation. Sentinel-2A and Landsat-9 images offer high spatial resolution, providing clearer patterns for distinct feature interpretation. ASTER's notable high spectral resolution compliments Sentinel-2A and Landsat-9 by projecting intricate feature spectral signatures, thereby enhancing the efficient discrimination of varied lithology in the scene.

Within the multisource data fusion framework [60], the ALOS PALSAR DEM data provide topographic insight which aids in the discrimination of surficial geological features in the scene. The gravity anomaly data projected the geophysical properties of subsurficial geological features. Integrating varied data in the multisource data fusion process enhances the spatial resolution of the low-resolution bands of images, as shown in Figure 10. It is observed in Figure 10 that the 30 m resolution of Landsat-9 and ASTER's band 5 is improved to 10 m with enriched textural information, thereby projecting feature details with better clarity.

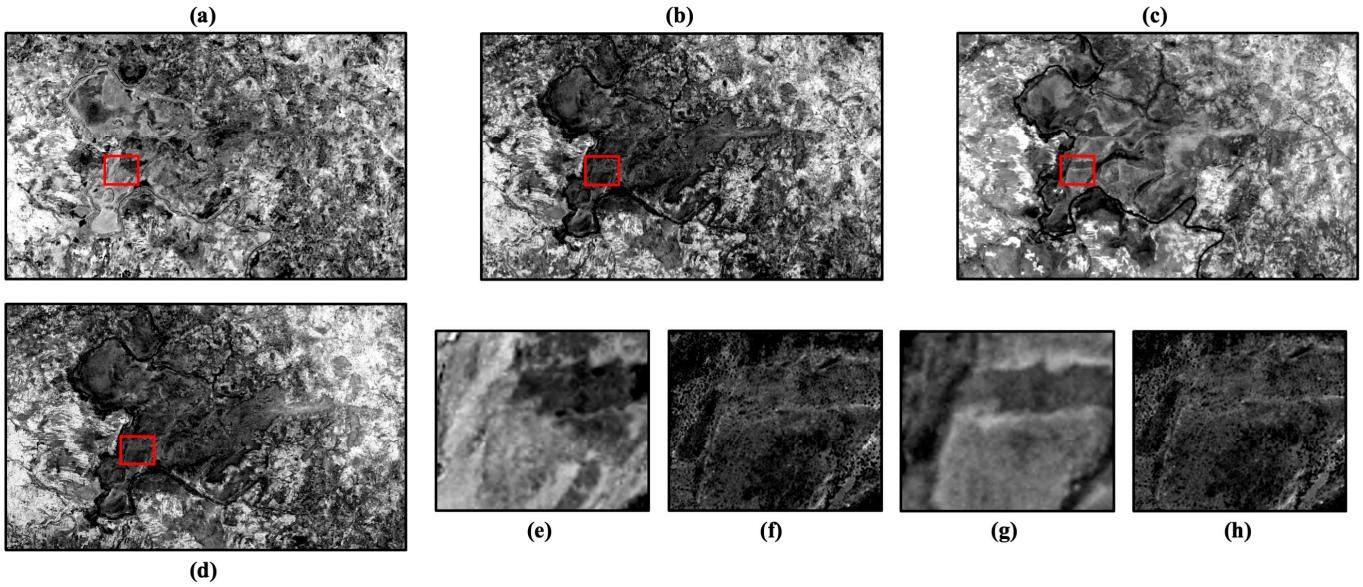

Figure 10. Band imagery: (**a**) Landsat-9 band 5; (**b**) Sentinel-2A band 5; (**c**) ASTER band 5; (**d**) fused image; (**e**) partial magnification of (**a**) (279 × 235 pixels); (**f**) partial magnification of (**b**) (279 × 235 pixels); (**g**) partial magnification of (**c**) (279 × 235 pixels); and (**h**) partial magnification of (**d**) (279 × 235 pixels).

The resultant fusion image from the multisource data, as shown in Figure 11, negates the limitations of the individual datasets. The synthesized fused imagery not only provides improved spatial resolution but also captures immense geophysical, spectral, and elevation information for a better characterization of the target lithology and the interpretation of representative features.

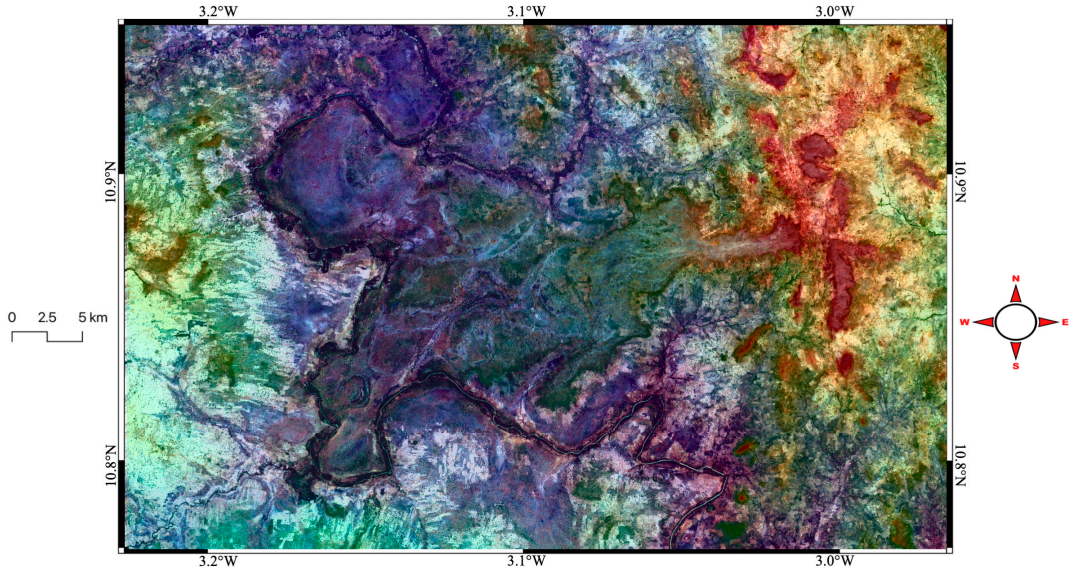

Figure 11. Resultant multisource fusion imagery.

5.3. Data Annotation and Sampling

The annotation map, serving as an essential feature in this study's geological analysis, is highlighted in Figure 12. The generation of the annotation map was based on the analysis of field data, geological feature interpretation from remote sensing image analysis, and utilizing geological resources [44,61]. These geological resources include the geological map of Burkina Faso published by SEMS Exploration Services [61]. With lithology being central to this study, the different rock types in the study area and their characteristics

were observed and defined. By conducting extensive fieldwork and engaging in expert geological analysis, six lithological classes that denote the geological feature representation in the study area were established. Furthermore, the presence of water bodies such as lakes, rivers, streams, and irrigation dams in the terrane was considered and clustered in a composite class as "water bodies" to alleviate probable interference with the overall classification process. This not only facilitates the precise labeling of classes but also ensures a distinctive comprehension of the varied geological descriptions in the terrane. The annotation map's dimensions in width and height were resampled to the size of the multisource fusion imagery, as shown in Figure 11.

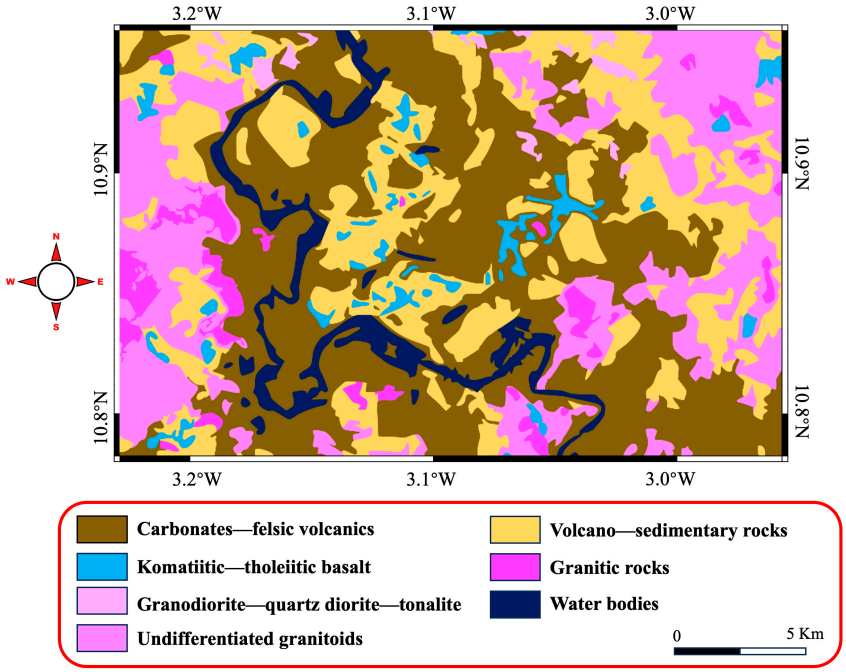

Figure 12. Annotation map of the study area.

The sample dataset for this study with spatial features was generated by integrating fieldwork and the sliding window algorithm on the multisource data at a stride of 2 (Figure 13). The configuration of the network structure in the sliding window and the extraction of discriminative spatial features are determined by the size of the window. The optimum unit size for the detection of lithological units was determined to be 3×3 after comparing the detection with varied window sizes; hence, the window size was set to 3×3. Implementing the sliding window method on the synthesized multisource fusion imagery generated 42,729 samples, which is a fraction of the total patches available. Each sample contained one or more labeled lithological classes.

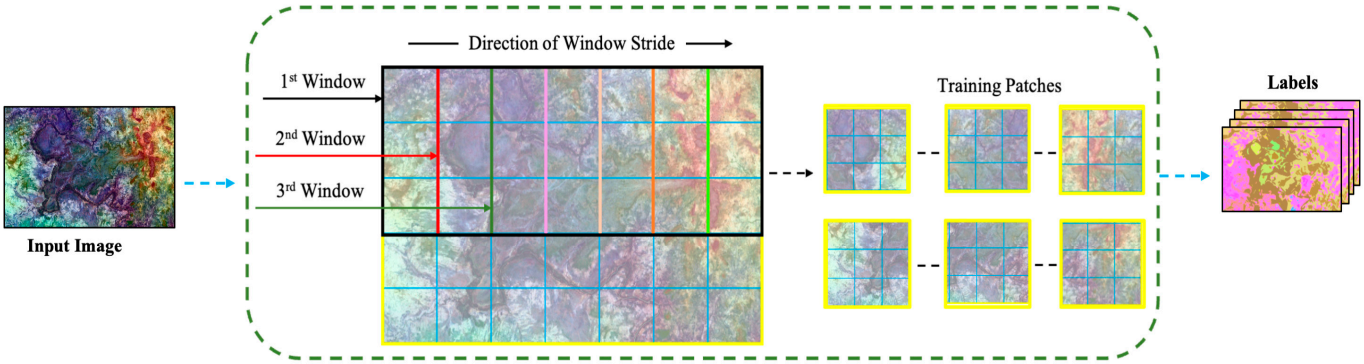

Figure 13. An illustration of the sliding window method implementation.

With ML being data-driven, the input dataset for the HAM-LSTM model was meticulously divided into training and test inputs to map lithological units in the fused multisource scene (Table 1). Constituting 70% of the input data, the training input serves in this study as a prerequisite to design the model structure and deduce the weights of variables and the bias matrix. The 30% test input was then used to evaluate HAM-LSTM's capability to generalize and forecast lithological features in the multisource fusion imagery.

Table 1. Outline of lithological categories of this study.

Lithological Units	Number of Training Samples (Pixel)	Number of Test Samples (Pixel)
Granitic rocks	6462	1949
Komatiitic–tholeiitic basalt	2165	1875
Water bodies	2170	1876
Undifferentiated granitoids	3235	2142
Carbonates–felsic volcanics	6796	2032
Volcano-sedimentary	4423	1439
Granodiorite–tonalite–quartz diorite	4661	1504
Total pixel number	29,912	12,817

5.4. Mapping Lithology with Hybrid Attention and LSTM Network

The HAM-LSTM model was trained and evaluated in the experimental design. Separate images were used for model training and testing to avoid overfitting. The training accuracy and loss curves of HAM-LSTM and the comparative model implementations are presented in Figure 14.

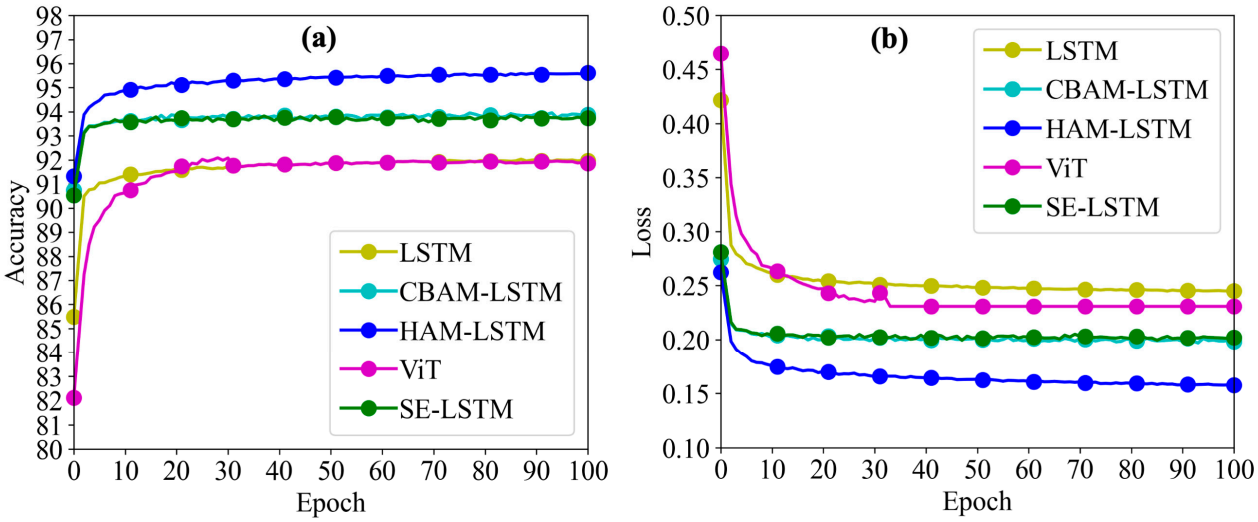

Figure 14. Graphs of training performance of the varied model implementations in this study: (**a**) accuracy and (**b**) loss.

The training metrics and convergence curves reveal the minimum loss and maximum training accuracy of HAM-LSTM after 100 training epochs, which were 0.16 and 95.56, respectively. This suggests that HAM-LSTM was efficiently trained and learned the underlying intricacies of the training data remarkably. Figure 15a highlights the classification map generated from the HAM-LSTM implementation. In view of the classification map, the lithological units identified by HAM-LSTM are in agreement with the annotation map and regional geological map, denoting consistency. However, a little "salt-and-pepper" effect can be observed. The classification map highlights the lithological composition of the study area as demonstrating features of a greenstone belt, which is notable for its geological implication.

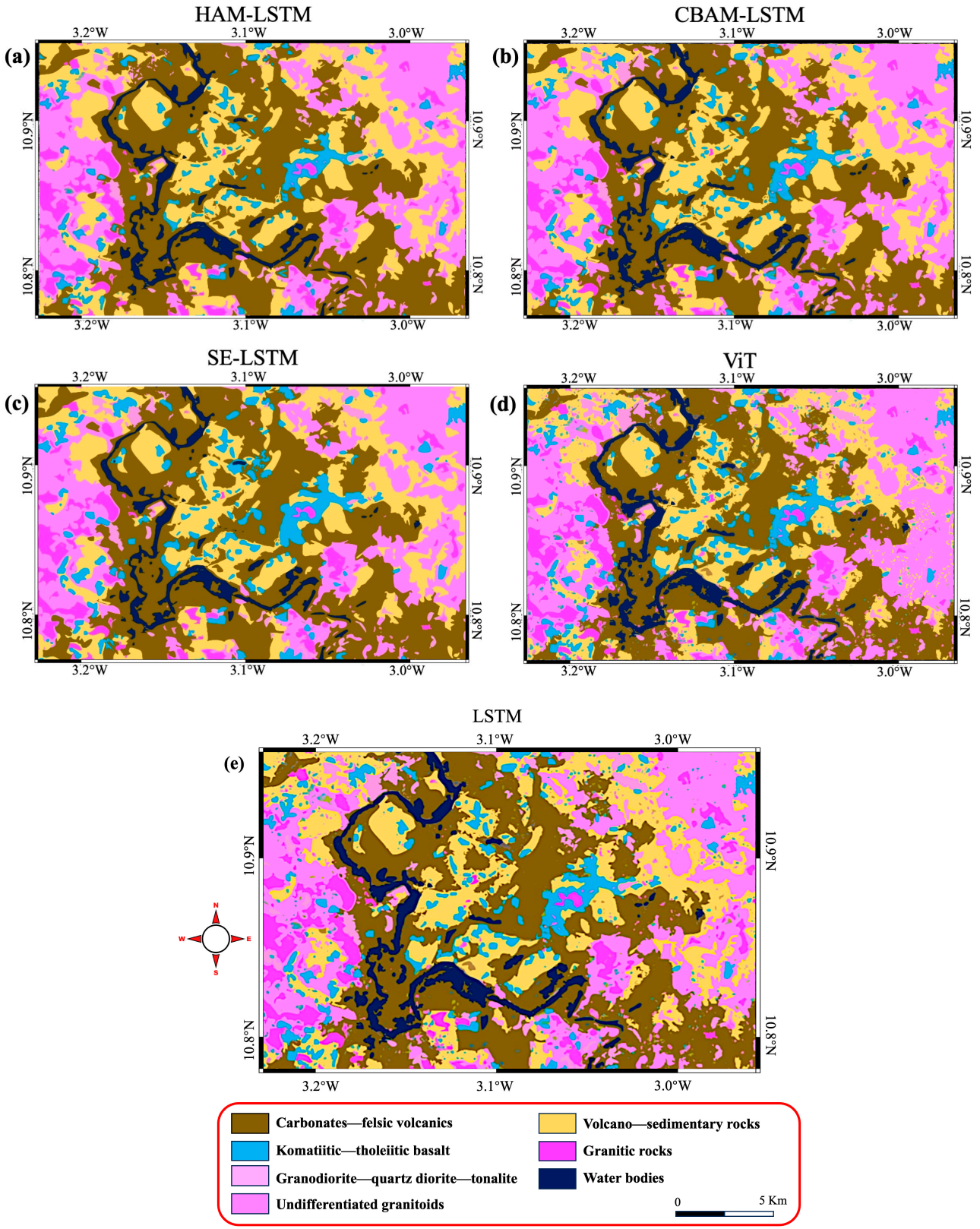

Figure 15. Classification maps derived from implementing (**a**) HAM-LSTM, (**b**) CBAM-LSTM, (**c**) SE-LSTM, (**d**) ViT, and (**e**) LSTM on the multisource fusion dataset.

Analyzing the output maps of the experiment projects intrusive rocks as the predominant lithological components in the scene, flanked by the sporadic distribution of komatiitic–tholeiitic basalts and the intercalation of carbonates and felsic volcanics. Insights

about the diverse geological heritage which includes magmatic intrusions, sedimentation, and volcanic activities of the study area are projected by the findings from this analysis, thereby accentuating the substantial metallogenic potential of the study area. The classification map does not only reveal the geological intricacy but also projects the terrane as significant for exploration and continued study in understanding geological processes.

5.5. Model Evaluation

The HAM-LSTM model performance and the comparative experiments were evaluated using the F1 score, overall accuracy (OA), precision, and recall. Computing the weighted-average precision and recall values of the models deduces the F1 score. Additionally, mean average precision (mAP) and confusion matrices were utilized to depict the classification performance of the model implementations. In deducing the evaluation metrics, the following equations were utilized:

$$Precision = \frac{TP}{FP + TP} \tag{36}$$

$$Recall = \frac{TP}{TP + FN} \tag{37}$$

$$F_1 - Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{38}$$

$$Overall\ Accuracy\ (OA) = \frac{TP + TN}{TP + FN + FP + TN} \tag{39}$$

$$Mean\ Average\ Precision\ (mAP) = \frac{1}{n}\sum_{k=1}^{k=n} AP_k \tag{40}$$

In the above Equations (36) to (39), the true positive, true negative, false positive, and false negative values are denoted by TP, TN, FP, and FN, respectively. In Equation (40), n denotes the number of classes, whereas the average precision for the kth class is denoted by AP_k.

Table 2 presents the result metrics, whereas Figure 16 highlights the confusion matrices of HAM-LSTM and the comparative implementations in the lithological classification scheme. From the result metrics, HAM-LSTM demonstrates a good generalization capability in classifying lithological units in the fused multisource imagery with an overall accuracy of 95.73%. The model shows a high precision in the recognition of lithological units consistent across all the lithological classes with an mAP of 0.96. The model's best recognition of lithological classes is observed in the undifferentiated granitoids with 0.98 precision, with the granitic rock class's 0.92 being the model's lowest precision in recognizing the lithological units.

Table 2. Classification result metrics of experimental sets used in this study.

Experiment	Class Name	Carbonates–Felsic Volcanics	Komatiitic–Tholeiitic Basalt	Granodiorite–Tonalite–Quartz Diorite	Undifferentiated Granitoids	Volcano-Sedimentary	Granitic Rock	Water Bodies
	Class ID	0	1	2	3	4	5	6
	Support	2032	1875	1504	2142	1439	1949	1876
HAM-LSTM	**Precision**	0.9523	0.9693	0.9653	0.9848	0.9820	0.9282	0.9835
	Recall	0.9388	0.8474	0.9688	0.9831	0.9835	0.9631	0.9984
	F1 Score	0.9455	0.9043	0.9671	0.9839	0.9726	0.9453	0.9909
				OA:	**0.9576**			
				mAP:	**0.9579**			
CBAM-LSTM	**Precision**	0.9532	0.8851	0.9633	0.9527	0.9447	0.9223	0.9981
	Recall	0.9126	0.8832	0.9625	0.9604	0.9812	0.9533	0.9467
	F1 Score	0.9324	0.8842	0.9629	0.9565	0.9626	0.9376	0.9717
				OA:	0.9447			
				mAP:	0.9452			

Table 2. *Cont.*

Experiment	Class Name	Carbonates–Felsic Volcanics	Komatiitic–Tholeiitic Basalt	Granodiorite–Tonalite–Quartz Diorite	Undifferentiated Granitoids	Volcano-Sedimentary	Granitic Rock	Water Bodies
SE-LSTM	**Precision**	0.9139	0.9391	0.9618	0.9863	0.9437	0.9116	0.9548
	Recall	0.9361	0.8678	0.9224	0.9606	0.9856	0.9095	0.9992
	F1 Score	0.9249	0.9021	0.9417	0.9733	0.9642	0.9105	0.9765
				OA:	0.9380			
				mAP:	0.9383			
ViT	**Precision**	0.9435	0.9481	0.9129	0.9637	0.9093	0.8904	0.9319
	Recall	0.9062	0.8535	0.9182	0.9209	0.9546	0.9228	0.9969
	F1 Score	0.9244	0.8983	0.9156	0.9418	0.9314	0.9063	0.9633
				OA:	0.9233			
				mAP:	0.9241			
LSTM	**Precision**	0.8902	0.8734	0.8429	0.9825	0.9646	0.9160	0.9965
	Recall	0.9226	0.7922	0.7112	0.9893	0.9910	0.9713	0.9925
	F1 Score	0.9061	0.8308	0.7715	0.9859	0.9777	0.9429	0.9945
				OA:	0.9207			
				mAP:	0.9193			

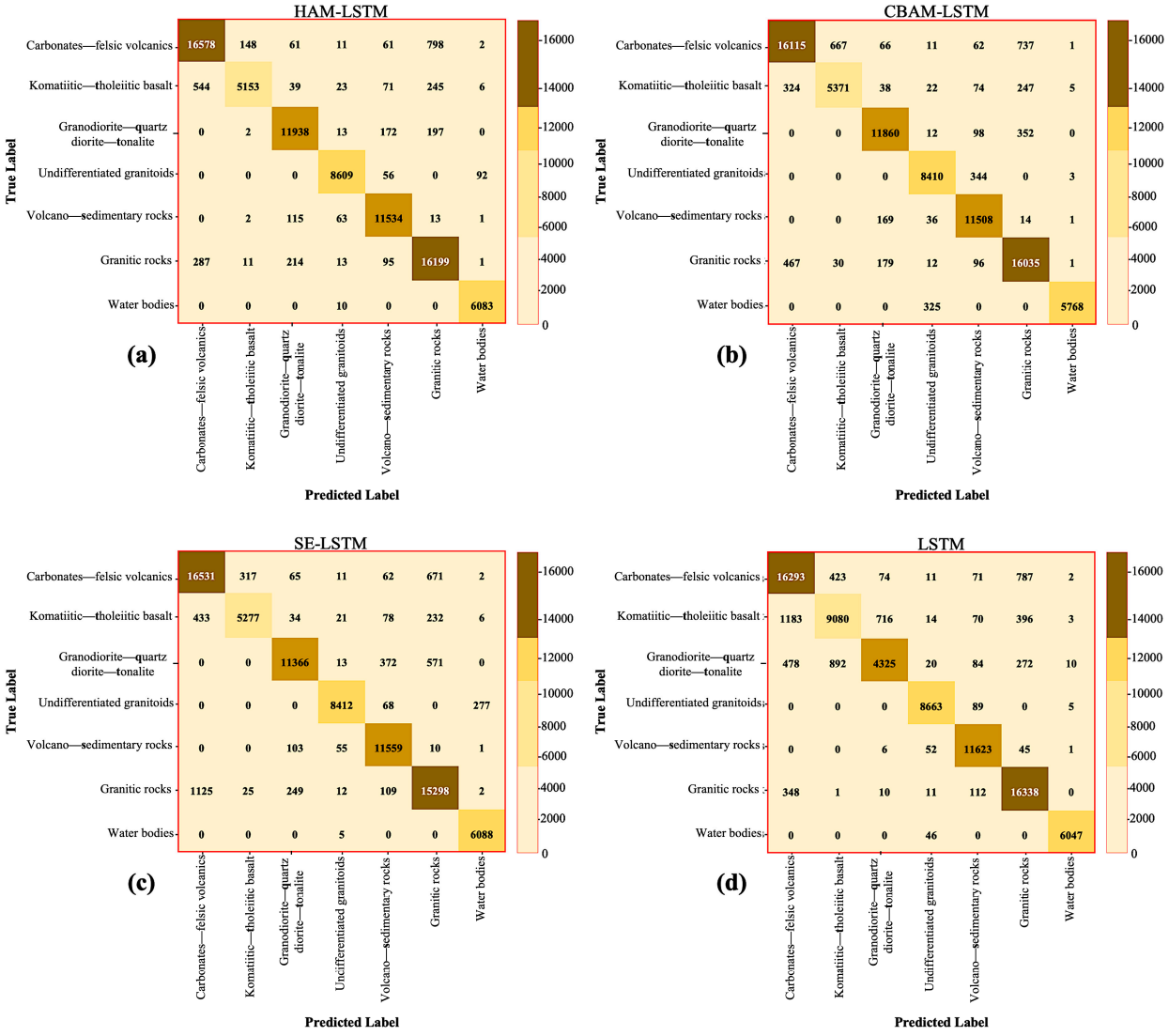

Figure 16. *Cont.*

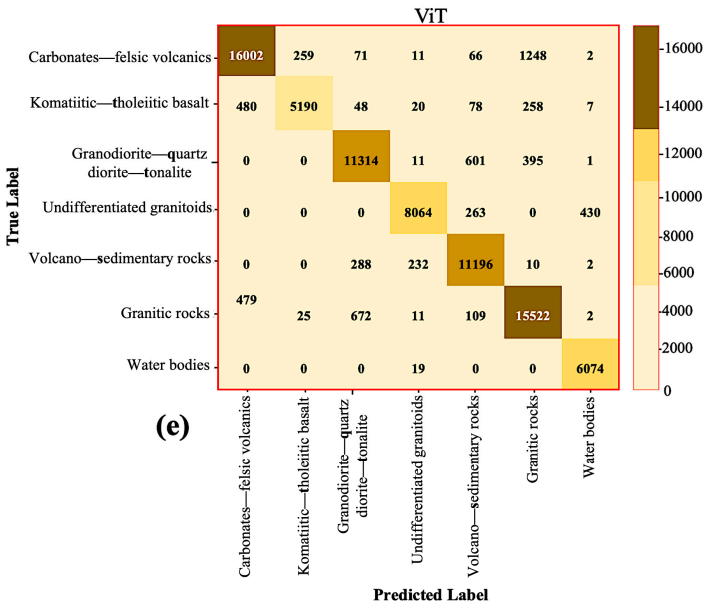

Figure 16. Confusion matrices of (**a**) HAM-LSTM, (**b**) CBAM-LSTM, (**c**) SE-LSTM, (**d**) LSTM, and (**e**) ViT implementation.

As shown in Figure 16a, the confusion matrix highlights HAM-LSTM's improved classification performance in all the lithological classes. This demonstrates that valued information from the five data sources, Sentinel-2A, Landsat-9, ASTER, ALOS PALSAR DEM, and gravity anomaly datasets, was effectively retained in the multisource data fusion process.

5.6. Comparative Experiments and Analysis

For a comprehensive evaluation of the HAM-LSTM model in the multisource lithological classification scheme, comparative experiments were conducted and analyzed with two other popular attention modules and a transformer model, namely, CBAM [62], SENet [27], and vision transformer (ViT) [63]. The CBAM and SENet modules were embedded into an LSTM framework to synthesize CBAM-LSTM and SE-LSTM, respectively, to classify lithological units in the multisource fusion dataset. ViT was selected as a candidate for the comparative analysis based on the multihead self-attention (MSA) module embedded in its transformer encoder. The models used in this comparative study were trained for 100 epochs on the training dataset and subsequently implemented on the multisource fusion test dataset. The evaluation metrics for this study comprised the recall, overall accuracy (OA), precision, F1 score, and confusion matrix. The mean average precision (mAP) of the models was also computed based on the precision values of the models in individual classes.

The experimental results are presented in Table 2, whereas Figures 15 and 16 highlight the classification maps and confusion matrices, respectively. The results in Table 2 and Figure 16 show HAM-LSTM's ability to attain state-of-the-art performance in the multisource lithological mapping scheme. The remarkable 95.76% OA of HAM-LSTM outperforms the runner-up model, CBAM-LSTM, by 1.29%. The 95.79% mAP of HAM-LSTM displaces SE-LSTM and ViT by 1.96% and 3.38%, respectively. From the ablation results in Table 3, HAM-LSTM improves the baseline LSTM model's OA by 3.69% and its mAP by 3.86% in multisource lithological classification.

Table 3. Result metrics of ablation study.

Method	LSTM (Base Model)	LSTM (Base Model)	LSTM (Base Model)	LSTM (Base Model)	LSTM (Base Model)	LSTM (Base Model)	LSTM (Base Model)	LSTM (Base Model)	HAM-LSTM	HAM-LSTM
Landsat-9	√									
ASTER		√								
Sentinel-2A			√							
ASTER + Landsat-9				√						
Sentinel-2A + Landsat-9					√					
ASTER + Sentinel-2A						√				
ASTER + Sentinel-2A + Landsat-9 + DEM							√			
Multisource								√	√	√
Channel attention + spatial attention									√	√
Channel separation										√
Overall accuracy (OA)/%	78.17	83.12	84.85	87.63	86.66	87.91	89.51	92.07	95.37	**95.76**
Mean average precision (mAP)	0.7888	0.8309	0.8455	0.8780	0.8663	0.8799	0.8955	0.9193	0.9542	**0.9579**

The HAM-LSTM model compared with the baseline LSTM, SE-LSTM, and CBAM is slightly complex in architecture and takes a relatively longer time to run. Notably, ViT had the highest computational density compared with all the models in the comparative experiment. It took the longest to run and yet had the lowest performance. HAM-LSTM's superior performance indicates that the size of a model, and the time it takes to run, does not solely determine a model's diagnostic, interpretation, and classification capabilities.

From the classification maps in Figure 15, all the models effectively identified lithological units in consonance with the annotation map and regional geological map, thereby proving consistency. However, a small "salt-and-pepper" effect can be observed in the classification maps of all the models. Despite CBAM-LSTM, SE-LSTM, and ViT having a small "salt-and-pepper" effect in classifying the lithological units, HAM-LSTM had better accuracy in classifying lithological units in the experiment.

The experimental analysis demonstrates that the proposed HAM-LSTM comparatively has a considerable edge over existing models in the diagnosis, interpretation, and classification of lithological units in remote sensing-based multisource geological interpretation. HAM-LSTM's remarkable performance makes it suitable for different industrial applications and continued research development.

5.7. Ablation Study

To assess the efficiency of the varied components of the HAM-LSTM model, an ablation study has been conducted in this study. With LSTM being the basis on which the proposed HAM-LSTM model is developed, LSTM is configured as the baseline model in the experimental design. The results from this experiment are presented in Table 3. Table 3 indicates that the baseline model implementation on Landsat-9 in classifying lithological units in the study area attained 78.17% OA with 78.88% mAP. Integrating the baseline configuration with ASTER attained 83.12% OA and 83.09% mAP. Sentinel-2A performed better with 84.85% OA and 84.55% mAP when integrated with the baseline model, averaging 4.21% OA and 3.57% mAP over Landsat-9 and ASTER. This can be attributed to the characteristic tonal information-capturing capacity of Sentinel-2A, which enhances intricate geological feature diagnosis and interpretation.

To quantify the advantages of integrating multisource data in remote sensing geological interpretation, experiments were conducted to compare different combinations of the fused data sources in the ablation study. The data source combination experiments included four varied data combinations, abbreviated as ASTER + Landsat-9, Sentinel-2A + Landsat-9, ASTER + Sentinel-2A, and Landsat-9 + ASTER + Sentinel-2A + ALOS PALSAR DEM, and a fifth combination involving all five sources of data. As Table 3 presents, the first four corresponding fused data combinations, ASTER + Landsat-9, Sentinel-2A + Landsat-9, ASTER + Sentinel-2A, and Landsat-9 + ASTER + Sentinel-2A + ALOS PALSAR DEM, recorded a significant increase in the overall accuracy at 87.63%, 86.66%,

87.91%, and 89.51%, respectively. The model achieved a 92.07% OA and 91.93% mAP when integrated with all five sources of fusion data comprising Setinel-2A, Landsat-9, ASTER, ALOS PALSAR DEM, and Bouguer anomaly gravity data. This improved performance proves that integrating multisource data in a remote sensing-based geological analysis not only provides more information but also enhances feature detection, diagnosis, and interpretation, relative to single-source data.

Subsequently, integrating the channel attention and spatial attention submodules of HAM into the model improved the model performance: the OA by 3.3% and the mAP by 3.45%. To complete the HAM-LSTM model design, the channel separation component was finally added to the model to partition the refined features' post-channel attention into two channel clusters along the channel dimension before implementing spatial attention. Integrating the HAM-LSTM model with the multisource fusion dataset further enhanced the model performance, which climaxed at 95.76% OA and 95.79% mAP.

The ablation experiment outlined and the results presented show the effect of the integrated multisource datasets, channel attention, spatial attention, and channel separation mechanisms on the performance of the HAM-LSTM model.

5.8. Complexity Analysis

In this study, key evaluation metrics such as the model parameter count and the giga-floating point operations per second (GFLOPs) were used to assess the computational requirements and complexity of the HAM module on the LSTM network against the squeeze-and-excitation network (SENet) and convolutional block attention mechanism (CBAM). Additionally, the training time for the attention-based models to complete an epoch was recorded. Table 4 presents the results of this analysis.

The experimental results indicate that the input size of the multisource fusion imagery used in the experimental set was uniform for all the models. Therefore, the GFLOPs of the competing modules were not influenced by the input size. With a $256 \times 256 \times 3$ input size, SE-LSTM had the shortest training time of 382 s, whereas HAM-LSTM required the longest training time at 630 s. As Table 4 depicts, CBAM-LSTM and SE-LSTM had the highest parameter count, surpassing HAM-LSTM by 0.14 M. This suggests that the overall parameter count of the HAM-LSTM model is not significantly impacted by the channel and spatial attention blocks and the integration of the channel separation operation within the HAM module framework. Despite the models having relatively similar parameter counts, with CBAM-LSTM and SE-LSTM having a slight edge over HAM-LSTM, the computational complexity of the models varied. CBAM-LSTM had the highest GFLOPs in comparison with SE-LSTM and HAM-LSTM. HAM-LSTM's GFLOPs were four times lower than those of CBAM-LSTM, indicating that the channel and spatial attention submodules and the HAM module framework had little effect on the computational complexity of the HAM-LSTM model.

The computational overhead reveals that the HAM-LSTM model not only supports real-time implementation but also has a high capacity to learn from data.

Table 4. Parameter count and computational complexity of attention-based LSTM models.

Model	Input Size	Training Time (s)	Parameters (M)	Computational Complexity (GFLOPs)
HAM-LSTM	$256 \times 256 \times 3$	630	29.40	0.03
SE-LSTM	$256 \times 256 \times 3$	382	29.54	0.05
CBAM-LSTM	$256 \times 256 \times 3$	489	29.54	0.12

6. Discussion

In this study, the quality of optimized recurrent neural networks (RNNs) in a proposed remote sensing experimental design is evaluated by employing qualitative and

quantitative assessments, as outlined in Sections 3–5. A high-performance improved RNN model, HAM-LSTM, based on a hybrid attention mechanism and LSTM is proposed. For a comprehensive performance evaluation, the HAM-LSTM model was implemented in a lithological classification scheme using multisource data comprising Setinel-2A, Landsat-9, ASTER, ALOS PALSAR DEM, and Bouguer anomaly gravity data. The classification results from the proposed model in the experimental design were then compared to those of other attention modules and deep learning models (SENet [27], CBAM [62], ViT [63], and LSTM [14]) in remote sensing geological interpretation using multisource data.

The results from the experiments, detailed in Figure 16 and Table 2, show that the HAM-LSTM model had the highest overall accuracy (OA) and mean average precision (mAP) of 95.76% and 0.96, respectively, outperforming the competing attention modules and attention-based models. However, HAM-LSTM showed some inconsistency in distinguishing individual lithological classes. Although the model averaged a 96.65% precision score in distinguishing individual lithological classes, it only scored the highest precision in distinguishing komatiitic–tholeiitic basalt, granodiorite–tonalite–quartz diorite, volcano-sedimentary, and granitic rock lithological classes. HAM-LSTM, thus, trailed CBAM-LSTM and SE-LSTM in distinguishing carbonates–felsic volcanics and undifferentiated granitoids at 95.32% and 98.63% precision, respectively. The findings from the experiments demonstrate the HAM-LSTM model's state-of-the-art performance in remote sensing geological interpretation.

Furthermore, visual analysis of the output classification maps in Figure 15 revealed that HAM-LSTM exhibited better homogeneity and minimized noise. Notably, HAM-LSTM showed more consistency in establishing lithological boundaries and a smoother appearance in comparison with the other competing models in the experimental sets. This suggests that the vegetation masking approach employed in this study created a balance in preserving detail and provided clarity for onward geological feature analysis and interpretation.

A significant merit of HAM-LSTM lies in the model's ability to comprehensively evaluate spectral similarities within an input dataset in its sequential iterative processing of individual pixels along the spectral axis. Also, the model can directly integrate the spatial context in feature learning to achieve appreciable accuracy in a lithological classification task. That notwithstanding, some limitations can be observed in HAM-LSTM's implementation. The HAM-LSTM model is relatively denser with a more complex model structure when compared with the baseline LSTM. During implementation, the HAM-LSTM model takes a relatively longer time to run than the baseline LSTM. Despite the consistency in identifying lithological classes with the annotation map of the study area, the output classification map of the HAM-LSTM model shows some "salt-and-pepper" effects. In future works, special attention will be channeled to the density and complexity of DL models with a focus on how to make improved models faster than their baseline variants for remote sensing geological interpretation.

6.1. Resource Use and Preprocessing Efficiency

Using the computer hardware and software specifications highlighted in Section 3.3, the preprocessing of the datasets was conducted with a scalability consideration for large-scale applications of the experimental design [7,18]. The radiometric calibration and atmospheric correction of the Landsat-9 and Sentinel-2A images were executed in thirty minutes, with each image taking approximately fifteen minutes. Executing atmospheric transformation, radiation correction, and orthophoto transformation on the ASTER product took about twenty minutes. The preprocessing of the gravity anomaly data and the synthesis of the complete Bouguer anomaly (CBA) and residual gravity maps took approximately thirty-five minutes, with the gaussian filtering requiring more resource intensity given the need for high-resolution adjustments.

The overall preprocessing of the datasets on the hardware and software resources used required about three and a half hours of computation and processing. This included the

preprocessing of the ALOS PALSAR DEM data and the process of masking the vegetation in the varied remote sensing images. The workflow, while intensive, was feasible on the hardware and software resources used in this study, making it suitable for large-scale applications and high-performance systems.

6.2. Geophysical Insights into Lithological Classification

Integrating Bouguer anomaly gravity data with multisensor remote sensing data is beneficial for geological analysis and interpretation [64,65]. Synthesizing the complete bouguer anomaly (CBA) and residual gravity maps from the Bouguer anomaly gravity data, as shown in Figure 9, provides critical insight into the study area's subsurface geological framework. The distinctive details such as the density variations in intrinsic lithological formations that gravity anomaly maps reveal make gravity anomaly data a valuable tool in remote sensing-based geological analysis and interpretation. In this study, the CBA and residual anomaly maps were integrated with multispectral remote sensing images to offer insights into identifying lithological units by delineating the gravity signatures of lithological feature representations, distinguishing subsurficial structural characteristics, and differentiating between regional and local geophysical influences on the area of interest.

The effect of regional gravity anomalies is minimized by the residual gravity map, as it highlights local geological features with improved clarity. This allows for the characteristic near-surface density variations in varied lithological units in the scene to be in focus, thereby projecting distinctive details about near-surface geological formations. For instance, the higher density of mafic, iron-rich units reflected by the high gravity signatures ranging between 1.1 mGal and 2.1 mGal are associated with Birimian metavolcanics. These gravity signatures align well with the known density contrasts and physical properties of the Birimian metavolcanics [65], thus underscoring the essence of their density contrast as distinguishable units in the area of interest. Relative to the residual gravity map, regional gravity anomalies are highlighted by the complete Bouguer anomaly (CBA) map, reflecting the effect of regional crustal and tectonic activities on deeper subsurface geological formations. Despite the potential of the effect of these regional influences masking the signatures of local geological features, the insights provided by the CBA map contextualize the effect of regional influences and make it easier to comprehend the overarching significance of deeper geological features in the area of interest. The characteristic low to moderate gravity responses of metasediments ranging from 0.2 mGal to 1.1 mGal are delineated in the residual gravity map given the diminished influence of regional gravity anomalies. The significance of gravity data in establishing lithological boundaries is reiterated by this intrinsic density contrast between the metavolcanics and metasediments in the area of interest.

In this study's multisource lithological classification scheme, geological interpretation is enhanced by the synergy between the multispectral remote sensing images and the Bouguer anomaly gravity data. This is a result of the integration of varied information including spectral and spatial detail, density contrast, and distinctive physical properties from the remote sensing and geophysical datasets, offering insights into surficial, near-surface, and deeper subsurface features. This multisource approach not only distinguishes lithological boundaries with better efficiency but also supports the identification of structural features that are crucial in delineating pathways of fluid flow in local mineralization processes. Also, the ability to extract distinctive details about the effect of regional and local anomalous geophysical signatures provides a better understanding of the underlying geological formation in a terrane when integrated with other multi-modal datasets. This sets the foundation for direct applications in crustal and tectonic studies and mineral exploration.

6.3. Impact of Attention Mechanism on Remote Sensing Geological Interpretation

Remote sensing provides an efficient roadmap in geological analysis and interpretation. However, some challenges arise when remote sensing techniques are integrated with

machine learning in geological surveys [50,66–69]. This becomes evident in comparison with manual interpretations due to factors such as irregular spatial distribution, sample imbalance, interclass resemblances, regolith, and geochemical similarities. These factors impede the accurate diagnosis, interpretation, and identification of geological features across varied remote sensing datasets. Integrating attention mechanisms in deep learning implementations in the context of remote sensing geological analysis has become essential for model improvement. This is based on the ability of attention modules to efficiently distinguish and assign computational resources with an emphasis on prioritizing relevant information in an input dataset [22–24,70,71].

As mentioned earlier, three attention modules, the squeeze-and-excitation network (SENet) [27], convolution block attention mechanism (CBAM) [62], and hybrid attention mechanism (HAM) [34], and a self-attention module incorporated in a vision transformer (ViT) [63] were employed in this study's experiments and analysis. The distinct characteristics of these modules make them different from each other. Their differences include their architectural framework, their computational principles, their core concepts, the type of methodology used for feature extraction and learning, and the parameters by which they categorize feature representations in a dataset. Consequently, the output of these modules in deep learning implementations in the context of remote sensing geological analysis is influenced by input data type and quality.

As presented in Figures 15 and 16, and Tables 2 and 4, the attention-based models HAM-LSTM, CBAM-LSTM, SE-LSTM, and ViT reported high accuracies in the lithological classification scheme using multisource data. Despite the high performance scores, the overall accuracies of the attention-based models varied. The models were employed on the same multisource fusion dataset, which provided improved spatial resolution and immense geophysical, spectral, and elevation information for a better characterization of target lithology and the interpretation of representative features. Therefore, the high performance scores of the corresponding attention-based models can be attributed to the multisource fusion data used. However, the varied overall accuracies of the attention-based models in the lithological classification scheme can be attributed to the varied distinct characteristics of the incorporated attention modules in the input dataset.

In employing the attention-based models in this study, incorporating the SE module in the SE-LSTM model effectively enhanced features that are important to current tasks within the network. The SE module improved the model's feature extraction and learning by computing each channel's attention weight, leveraging global average pooling and fully connected layers. This was followed by the scaling of the features of each channel, thereby enhancing important channel features. However, the susceptibility of the SE module to rely on fully connected layers to compute channel attention weights within the model framework increased the model's computational overhead. Substituting the SE module in the LSTM network with the CBAM module resulted in an improvement in performance gain. Incorporating both channel attention and spatial attention, the CBAM module facilitated the capturing of a broader range of important features in the input multisource fusion imagery. The capability of CBAM to simultaneously extract and learn representative feature information in both channel and spatial dimensions makes the module a proficient tool for fine-grained and large-scale feature differentiation tasks. This includes tasks like identifying lithological units in remote sensing geological analysis. That notwithstanding, the need for additional submodules in a CBAM block within a model to compute channel and spatial attention weights comes at the cost of increased model computational overhead, just like the SE module.

Alternately, the HAM module improved feature learning by enhancing spectral and spatial feature integration. Within the HAM-LSTM framework, the HAM submodules were configured with an adaptive operation that combined the global-average-pooled features and max-pooled features. HAM effectively captured local interactions across channels within the input multisource fusion imagery, leveraging one-dimensional convolutions. With 1D convolutions, HAM regulated the range of the captured local interactions by

adjusting the kernel size automatically. In this configuration, comprehensive inter-channel information was preserved by the HAM module within the HAM-LSTM model framework, thereby curbing the need for dimensionality reduction operations and maintaining important feature dependencies within the network. Furthermore, a feature separation chamber with a channel separation operation within the module designates refined channel features from resultant channel attention maps into composite feature clusters by the channel dimension. Subsequently, feature interactions across spatial dimensions were captured by integrating a 7×7 convolution layer on feature descriptors to emphasize and enhance critical features. Despite the relative increase in processing time, this technique enables the HAM module to effectively capture cross-channel and cross-spatial interactions with fewer parameters and reduced computational overhead. Within the HAM-LSTM model framework, the HAM module offers a proficient lightweight solution for improved feature information capturing across channel and spatial dimensions without significantly impacting model complexity. The results from the experimental sets prove that the proposed HAM-LSTM model considerably improves lithological classification accuracy in remote sensing-based geological interpretation.

In contrast, the self-attention module, as part of ViT's transformer framework, distinguishes it from other attention-based models based on its computational principles and core concepts. For classification tasks, ViT relies on the global context between image patches in extracting distinctive generalized features [63]. However, challenges such as high-dimensional computations, limited parameter interdependency, and dense connections within the model framework make ViT susceptible to high data dependency, local extrema, irregular feature correlation, and overshooting [72–74]. This increases the model's complexity and computational overhead.

Overall, the findings from this study highlight the performance of varied attention modules in attention-based models. The performance of these models is influenced by input data composition and the architectural framework of the model. Attention modules such as HAM, SE, CBAM, and transformer-configured self-attention demonstrate significant performance gains with varied computational requirements. HAM, however, demonstrates proficiency in geological applications with low computational cost. It is therefore imperative to create a balance between efficiency and task-specific accuracy when choosing an attention module for remote sensing geological interpretation. This study reiterates the essence of attention modules in integrating machine learning with remote sensing for geological studies.

7. Conclusions

This study presents an efficient roadmap by integrating attention-based deep learning methods and multisource data fusion techniques for improved remote sensing geological interpretation. Leveraging multisource data fusion in the experimental design enriched the diagnostic information for enhanced feature discrimination. HAM-LSTM, an improved hybrid RNN framework that integrates a hybrid attention mechanism into a long short-term memory (LSTM) network, was developed to map lithological units, with a case study focused on parts of the West African Craton. From the experimental results, three key insights can be drawn:

(1) Integrating geophysical data such as the gravity anomaly data with multi-modal remote sensing datasets, including ASTER, Landsat-9, Sentinel-2A, and ALOS PALSAR DEM, provided more information and enhanced feature detection, diagnosis, and interpretation, relative to using a single-source data.

(2) The HAM-LSTM model incorporating channel and spatial attention with a channel separation operation excels in capturing local interactions across channels and spatial dimensions on random-sized inputs within the network. Addressing the limitation of the baseline LSTM, HAM-LSTM's integrated configuration fully captured the spectral and spatial interrelations amongst feature representations in the multisource fusion dataset, thereby improving geological interpretation.

(3) The results from extensive experiments have demonstrated that the proposed model achieves state-of-the-art performance and outperforms existing attention modules and attention-based models. Compared with the baseline LSTM, HAM-LSTM showed significant performance gains and identified lithological units, improving the accuracy by 3.69%.

This study's approach underscores the significance of leveraging deep learning and multisource datasets in remote sensing geological analysis. This sets the foundation for direct applications in geospatial analysis, crustal and tectonic studies, and mineral exploration. Further studies on integrating remote sensing technology with deep learning to map lithology will be conducted in future works to expand on insights drawn from this study and propose more efficient geological interpretation models using lightweight architectural frameworks.

Author Contributions: Conceptualization, M.A.-T. and W.X.; methodology, M.A.-T.; software, M.A.-T.; validation, M.A.-T., W.X. and E.D.S.; visualization, M.A.-T., W.X. and E.D.S.; formal analysis, M.A.-T.; investigation, M.A.-T.; resources, M.A.-T. and W.X.; data curation, M.A.-T. and W.X.; writing—original draft, M.A.-T.; writing–review and editing, M.A.-T. and E.D.S.; supervision, W.X.; project administration, M.A.-T. and W.X.; funding acquisition, W.X. All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported by the Science and Technology Program of Sichuan (grant number 2023ZHCG0013) and the local special funded project of the central government's Mianyang Municipality Science and Technology Bureau (grant number 2022ZYDF009).

Data Availability Statement: The datasets used in this study, including the remote sensing and gravity anomaly data, are available for free at the following platforms: the United States Geological Survey (https://earthexplorer.usgs.gov/); the European Space Agency's Copernicus Data Space Ecosystem (https://dataspace.copernicus.eu/); the Alaska Satellite Facility (https://asf.alaska.edu/datasets/daac/alos-palsar/); and the GFZ—German Research Centre for Geosciences (www.gfz-potsdam.de). Accessed on 28 October 2022.

Acknowledgments: This study is part of the Doctoral Research Fellowship of the first author under the guidance of the second author at the School of Resources and Environment of the University of Electronic Science and Technology of China, China. Therefore, the authors acknowledge the University of Electronic Science and Technology of China for granting the Doctoral Research Fellowship to the first author.

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

References

1. 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-Inform.* **2019**, *8*, 248. [CrossRef]
2. 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]
3. Shebl, A.; Hissen, M.; Abdelaziz, M.I.; Csámer, Á. Lithological mapping enhancement by integrating Sentinel 2 and gamma-ray data utilizing support vector machine: A case study from Egypt. *Int. J. Appl. Earth Obs. Geoinf.* **2021**, *105*, 102619. [CrossRef]
4. Liu, T.; Chen, T.; Niu, R.; Plaza, A.J. Landslide Detection Mapping Employing CNN, ResNet, and DenseNet in the Three Gorges Reservoir, China. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 11417–11428. [CrossRef]
5. 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]
6. Bolouki, S.M.; Ramazi, H.R.; Maghsoudi, A.; Beiranvand Pour, A.; Sohrabi, G. A Remote sensing-based application of bayesian networks for epithermal gold potential mapping in Ahar-Arasbaran area, NW Iran. *Remote Sens.* **2020**, *12*, 105. [CrossRef]
7. Wang, Z.; Zuo, R.; Liu, H. Lithological mapping based on fully convolutional network and multi-source geological data. *Remote Sens.* **2021**, *13*, 4860. [CrossRef]
8. Li, Y.; Chen, R.; Zhang, Y.; Zhang, M.; Chen, L. Multi-label remote sensing image scene classification by combining a convolutional neural network and a graph neural network. *Remote Sens.* **2020**, *12*, 4003. [CrossRef]

9. 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]
10. Jia, S.; Liao, J.; Xu, M.; Li, Y.; Zhu, J.; Sun, W.; Jia, X.; Li, Q. 3-D Gabor Convolutional Neural Network for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5509216. [CrossRef]
11. Wang, Z.; Li, T.; Zuo, R. Leucogranite mapping via convolutional recurrent neural networks and geochemical survey data in the Himalayan orogen. *Geosci. Front.* **2024**, *15*, 101715. [CrossRef]
12. Ouyang, S.; Ouyang, S.; Chen, W.; Chen, W.; Qin, X.; Qin, X.; Yang, J.; Yang, J. Geological Background Prototype Learning-Enhanced Network for Remote-Sensing-Based Engineering Geological Lithology Interpretation in Highly Vegetated Areas. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2024**, *17*, 8794–8809. [CrossRef]
13. Cho, K.; van Merriënboer, B.; Gulcehre, C.; Bahdanau, D.; Bougares, F.; Schwenk, H.; Bengio, Y. Learning Phrase Representations Using RNN Encoder-Decoder for Statistical Machine Translation. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Doha, Qatar, 25–29 October 2014; pp. 1724–1734. [CrossRef]
14. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef] [PubMed]
15. Mou, L.; Ghamisi, P.; Zhu, X.X. Deep recurrent neural networks for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 3639–3655. [CrossRef]
16. Zhang, X.; Sun, Y.; Jiang, K.; Li, C.; Jiao, L.; Zhou, H. Spatial Sequential Recurrent Neural Network for Hyperspectral Image Classification. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2018**, *11*, 4141–4155. [CrossRef]
17. Wu, H.; Prasad, S. Convolutional Recurrent Neural Networks for Hyperspectral Data Classification. *Remote Sens.* **2017**, *9*, 298. [CrossRef]
18. 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*, 4510314. [CrossRef]
19. Cheng, G.; Xie, X.; Han, J.; Guo, L.; Xia, G.-S. Remote Sensing Image Scene Classification Meets Deep Learning: Challenges, Methods, Benchmarks, and Opportunities. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *13*, 3735–3756. [CrossRef]
20. Cao, J.; Bao, W.; Shang, H.; Yuan, M.; Cheng, Q. GCL-YOLO: A GhostConv-Based Lightweight YOLO Network for UAV Small Object Detection. *Remote Sens.* **2023**, *15*, 4932. [CrossRef]
21. Liu, Q.; Zhou, F.; Hang, R.; Yuan, X. Bidirectional-Convolutional LSTM Based Spectral-Spatial Feature Learning for Hyperspectral Image Classification. *Remote Sens.* **2017**, *9*, 1330. [CrossRef]
22. Yang, X. An Overview of the Attention Mechanisms in Computer Vision. *J. Phys. Conf. Ser.* **2020**, *1693*, 012173. [CrossRef]
23. Niu, Z.; Zhong, G.; Yu, H. A review on the attention mechanism of deep learning. *Neurocomputing* **2021**, *452*, 48–62. [CrossRef]
24. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention Is All You Need. In Proceedings of the 31st Conference on Information Proceeding Systems, Long Beach, CA, USA, 4–9 December 2017. [CrossRef]
25. Zhu, H.; Xie, C.; Fei, Y.; Tao, H. Attention Mechanisms in CNN-Based Single Image Super-Resolution: A Brief Review and a New Perspective. *Electronics* **2021**, *10*, 1187. [CrossRef]
26. Qin, Z.; Zhang, P.; Wu, F.; Li, X. FcaNet: Frequency Channel Attention Networks. In Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada, 10–17 October 2021; pp. 763–772. [CrossRef]
27. Hu, J.; Shen, L.; Sun, G. Squeeze-and-Excitation Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–23 June 2018; pp. 7132–7141. [CrossRef]
28. Wang, Q.; Wu, B.; Zhu, P.; Li, P.; Zuo, W.; Hu, Q. ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 14–19 June 2020; pp. 11531–11539. [CrossRef]
29. Zhao, H.; Zhang, Y.; Liu, S.; Shi, J.; Loy, C.; Lin, D.; Jia, J. PSANet: Point-wise Spatial Attention Network for Scene Parsing. In Proceedings of the 2018 European Conference on Computer Vision, Munich, Germany, 8–14 September 2018; Available online: https://github.com/hszhao/PSANet (accessed on 15 April 2024).
30. Hu, J.; Shen, L.; Albanie, S.; Sun, G.; Vedaldi, A. Gather-Excite: Exploiting Feature Context in Convolutional Neural Networks. *Adv. Neural Inf. Process. Syst.* **2018**, *31*. [CrossRef]
31. Liu, Z.; Mao, H.; Wu, C.-Y.; Feichtenhofer, C.; Darrell, T.; Xie, S. A ConvNet for the 2020s. In Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 18–24 June 2022; pp. 11966–11976. [CrossRef]
32. Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang, Z.; Lin, S.; Guo, B. Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows. In Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada, 10–17 October 2021; pp. 9992–10002. [CrossRef]
33. Guo, M.-H.; Lu, C.-Z.; Liu, Z.-N.; Cheng, M.-M.; Hu, S.-M. Visual Attention Network. *Comput. Vis. Media* **2022**, *9*, 733–752. [CrossRef]
34. Li, G.; Fang, Q.; Zha, L.; Gao, X.; Zheng, N. HAM: Hybrid attention module in deep convolutional neural networks for image classification. *Pattern Recognit.* **2022**, *129*, 108785. [CrossRef]
35. Chen, X.; Wang, X.; Zhou, J.; Qiao, Y.; Dong, C. Activating More Pixels in Image Super-Resolution Transformer. In Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, BC, Canada, 17–24 June 2023; pp. 22367–22377. [CrossRef]

36. Luong, M.; Pham, H.; Manning, C.D. Effective Approaches to Attention-based Neural Machine Translation. In Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing (EMNLP), Lisbon, Portugal, 17–21 September 2015; pp. 1412–1421. [CrossRef]
37. Xu, K.; Ba, J.; Kiros, R.; Cho, K.; Courville, A.; Salakhudinov, R.; Zemel, R.; Bengio, Y. Show, Attend and Tell: Neural Image Caption Generation with Visual Attention. In Proceedings of the 32nd International Conference on Machine Learning, ICML 2015, Lille, France, 6–11 July 2015; Volume 3, pp. 2048–2057. Available online: https://arxiv.org/abs/1502.03044v3 (accessed on 7 May 2024).
38. Subuda, S.; Yao, H.; Bai, Y.; Li, J. Image Recognition Based on improved LSTM deep neural network. In Proceedings of the The International Conference on Forthcoming Networks and Sustainability (FoNeS 2022), Nicosia, Cyprus, 3–5 October 2022; pp. 17–21. [CrossRef]
39. He, Y.; Li, W.; Dong, Z.; Zhang, T.; Shi, Q.; Wang, L.; Wu, L.; Qian, S.; Wang, Z.; Liu, Z.; et al. Lithologic Identification of Complex Reservoir Based on PSO-LSTM-FCN Algorithm. *Energies* **2023**, *16*, 2135. [CrossRef]
40. Zhu, H.; Wang, X.; Chen, X.; Zhang, L. Similarity search and performance prediction of shield tunnels in operation through time series data mining. *Autom. Constr.* **2020**, *114*, 103178. [CrossRef]
41. Jessell, M.W.; Liégeois, J.-P. 100 years of research on the West African Craton. *J. Afr. Earth Sci.* **2015**, *112*, 377–381. [CrossRef]
42. Thiéblemont, D.; Delor, C.; Cocherie, A.; Lafon, J.M.; Goujou, J.C.; Baldé, A.; Bah, M.; Sané, H.; Fanning, C.M. A 3.5 Ga granite–gneiss basement in Guinea: Further evidence for early archean accretion within the West African Craton. *Precambrian Res.* **2001**, *108*, 179–194. [CrossRef]
43. Markwitz, V.; Hein, K.A.; Miller, J. Compilation of West African mineral deposits: Spatial distribution and mineral endowment. *Precambrian Res.* **2016**, *274*, 61–81. [CrossRef]
44. Bessoles, B. Géologie de l'Afrique. Vol 1 Le Craton Ouest Africain. *Bur. Rech. Géol. Min. Mém.* **1977**, *88*, 402–479.
45. Abrams, M.; Yamaguchi, Y. Twenty Years of ASTER Contributions to Lithologic Mapping and Mineral Exploration. *Remote Sens.* **2019**, *11*, 1394. [CrossRef]
46. Aeronautics, N.; Administration, S. Landsat 9 Mission Brochure. Available online: https://www.usgs (accessed on 1 December 2022).
47. Blakely, R.J.; Simpson, R.W. Approximating edges of source bodies from magnetic or gravity anomalies. *Geophysics* **1986**, *51*, 1494–1498. [CrossRef]
48. Blakely, R.J. *Potential Theory in Gravity and Magnetic Applications*; Cambridge University Press: Cambridge, UK, 1995. [CrossRef]
49. Rouse, J.W.; Haas, R.H.; Schell, J.A.; Deering, D.W. Monitoring Vegetation Systems in the Great Plains with ERTS. In *Third Earth Resources Technology Satellite-1 Symposium—Volume I: Technical Presentations, NASA SP-351*; NASA Special Publication: Washington, DC, USA, 1974; Volume 1, pp. 48–62.
50. 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]
51. Dempster, A.P. Upper and Lower Probabilities Induced by a Multivalued Mapping. *Ann. Math. Statist.* **1967**, *38*, 325–339. [CrossRef]
52. Shafer, G. *A Mathematical Theory of Evidence*; Princeton University Press: Princeton, NJ, USA, 1976.
53. Brown, D.E.; Duren, B.G. Conflicting information integration for decision support. *Decis. Support Syst.* **1986**, *2*, 321–329. [CrossRef]
54. Hall, D.; Llinas, J. An introduction to multisensor data fusion. *Proc. IEEE* **1997**, *85*, 6–23. [CrossRef]
55. Ding, H.; Jing, L.; Li, H.; Tang, Y.; Ma, H.; Zhu, B.; Wang, W.; Qiu, L. Method and System for Improving Resolution of Geochemical Element Layer. Chinese Patent no. 201811275285.4, 2018.
56. Adelson, E.H.; Anderson, C.H.; Bergen, J.R.; Burt, P.J.; Ogden, J.M. Pyramid methods in image processing. *RCA Eng.* **1984**, *29*, 33–41.
57. Aiazzi, B.; Alparone, L.; Barducci, A.; Baronti, S.; Pippi, I. Multispectral fusion of multisensor image data by the generalized Laplacian pyramid. In Proceedings of the IEEE 1999 International Geoscience and Remote Sensing Symposium. IGARSS'99, Hamburg, Germany, 28 June–2 July 1999; Volume 2, pp. 1183–1185.
58. Keys, R. Cubic Convolution Interpolation for Digital Image Processing. *IEEE Trans. Acoust.* **1981**, *29*, 1153–1160. [CrossRef]
59. Kingma, D.P.; Ba, J.L. Adam: A Method for Stochastic Optimization. In Proceedings of the 3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, 7–9 May 2015. Available online: https://arxiv.org/abs/1412.6980v9 (accessed on 7 May 2024).
60. Ding, H.; Jing, L.; Xi, M.; Bai, S.; Yao, C.; Li, L. Research on Scale Improvement of Geochemical Exploration Based on Remote Sensing Image Fusion. *Remote Sens.* **2023**, *15*, 1993. [CrossRef]
61. Home—SEMS Exploration Services. Available online: https://sems-exploration.com/ (accessed on 30 October 2024).
62. Woo, S.; Park, J.; Lee, J.-Y.; Kweon, I.S. *CBAM: Convolutional Block Attention Module*; Springer: Berlin/Heidelberg, Germany, 2018; pp. 3–19. [CrossRef]
63. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. An image is worth 16x16 words: Transformers for image recognition at scale. In Proceedings of the ICLR 2021—9th International Conference on Learning Representations, Virtual, 3–7 May 2021. [CrossRef]

64. Xiao, F.; Wang, Z. Geological interpretation of Bouguer gravity and aeromagnetic data from the Gobi-desert covered area, Eastern Tianshan, China: Implications for porphyry Cu-Mo polymetallic deposits exploration. *Ore Geol. Rev.* **2017**, *80*, 1042–1055. [CrossRef]
65. Forson, E.D.; Menyeh, A.; Wemegah, D.D. Mapping lithological units, structural lineaments and alteration zones in the Southern Kibi-Winneba belt of Ghana using integrated geophysical and remote sensing datasets. *Ore Geol. Rev.* **2021**, *137*, 104271. [CrossRef]
66. 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]
67. Maxwell, A.E.; Warner, T.A.; Fang, F. Implementation of machine-learning classification in remote sensing: An applied review. *Int. J. Remote Sens.* **2018**, *39*, 2784–2817. [CrossRef]
68. Shirmard, H.; Farahbakhsh, E.; Müller, R.D.; Chandra, R. A review of machine learning in processing remote sensing data for mineral exploration. *Remote Sens. Environ.* **2022**, *268*, 112750. [CrossRef]
69. Shirmard, H.; Farahbakhsh, E.; Heidari, E.; Pour, A.B.; Pradhan, B.; Müller, D.; Chandra, R. A Comparative Study of Convolutional Neural Networks and Conventional Machine Learning Models for Lithological Mapping Using Remote Sensing Data. *Remote Sens.* **2022**, *14*, 819. [CrossRef]
70. Mei, S.; Li, X.; Liu, X.; Cai, H.; Du, Q. Hyperspectral Image Classification Using Attention-Based Bidirectional Long Short-Term Memory Network. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5509612. [CrossRef]
71. Appiah-Twum, M.; Xu, W. An Attention-Based LSTM Lithological Classification Using Multisensor Datasets. In Proceedings of the IGARSS 2024-2024 IEEE International Geoscience and Remote Sensing Symposium, Athens, Greece, 7–12 July 2024; pp. 2136–2140. [CrossRef]
72. Dong, Y.; Cordonnier, J.B.; Loukas, A. Attention is Not All You Need: Pure Attention Loses Rank Doubly Exponentially with Depth. In Proceedings of the 38th International Conference on Machine Learning, Virtual, 18–24 July 2021. [CrossRef]
73. Chen, Z.; Xie, L.; Niu, J.; Liu, X.; Wei, L.; Tian, Q. Visformer: The Vision-Friendly Transformer. In Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada, 10–17 October 2021; pp. 569–578. [CrossRef]
74. Appiah-Twum, M.; Xu, W.; Acheampong, E.M. DenseViT: A Hybrid CNN-Vision Transformer Model for an Improved Multi-sensor Lithological Classification. In Proceedings of the IGARSS 2024-2024 IEEE International Geoscience and Remote Sensing Symposium, Athens, Greece, 7–12 July 2024; pp. 3418–3422. [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.