

Research Article

Deep Subsurface Pseudo-Lithostratigraphic Modeling Based on Three-Dimensional Convolutional Neural Network (3D CNN) Using Inversed Geophysical Properties and Shallow Subsurface Geological Model

Baoyi Zhang¹,² Zhanghao Xu,¹ Xiuzong Wei,² Lei Song,^{1,3} Syed Yasir Ali Shah,¹ Umair Khan^{1,4}, Linze Du,¹ and Xuefeng Li⁵

¹Key Laboratory of Metallogenetic Prediction of Nonferrous Metals and Geological Environment Monitoring (Ministry of Education), School of Geosciences and Info-Physics, Central South University, Changsha, 410083, China

²Geotechnical Engineering Investigation and Design Institute, Guangxi Communications Design Group Co. Ltd., Nanning, 530029, China

³Hunan Institute of Geological Disaster Investigation and Monitoring, Changsha, 410004, China

⁴Institute of Deep-Sea Science and Engineering, Chinese Academy of Sciences, Sanya, 572000, China

⁵Natural Resources Survey Institute of Heilongjiang Province, Harbin, 150094, China

Correspondence should be addressed to Umair Khan; umairkhanbku@mails.ucas.ac.cn

Received 8 September 2023; Accepted 18 January 2024; Published 16 February 2024

Academic Editor: Jinbo Chen

Copyright © 2024. Baoyi Zhang et al. Exclusive Licensee GeoScienceWorld. Distributed under a Creative Commons Attribution License (CC BY 4.0).

Lithostratigraphic modeling holds a vital role in mineral resource exploration and geological studies. In this study, we introduce a novel approach for automating pseudo-lithostratigraphic modeling in the deep subsurface, leveraging inversed geophysical properties. We propose a three-dimensional convolutional neural network with adaptive moment estimation (3D Adam-CNN) to achieve this objective. Our model employs 3D geophysical properties as input features for training, concurrently reconstructing a 3D geological model of the shallow subsurface for lithostratigraphic labeling purposes. To enhance the accuracy of pseudo-lithostratigraphic modeling during the model training phase, we redesign the 3D CNN framework, fine-tuning its parameters using the Adam optimizer. The Adam optimizer ensures controlled parameter updates with minimal memory overhead, rendering it particularly well-suited for convolutional learning involving huge 3D datasets with multi-dimensional features. To validate our proposed 3D Adam-CNN model, we compare the performance of our approach with 1D and 2D CNN models in the Qingniandian area of Heilongjiang Province, Northeastern China. By cross-matching the model's predictions with manually modeled shallow subsurface lithostratigraphic distributions, we substantiate its reliability and accuracy. The 3D Adam-CNN model emerges as a robust and effective solution for lithostratigraphic modeling in the deep subsurface, utilizing geophysical properties.

1. Introduction

Litho-strata manifest a broad range of rock properties arising from distinct geological processes, for example, weathering, compaction, metamorphism, and deformation. These intricate processes, when combined, yield complex representations of various litho-strata. Furthermore,

lithostratigraphy dictates the physical and chemical attributes of the litho-strata, the distribution of which is intimately intertwined with mineral resource distribution. Traditional lithostratigraphic identification methods, for example, borehole drilling and trenching, necessitate manual interpretation of field-collected or borehole core samples. While direct, these methods prove costly,

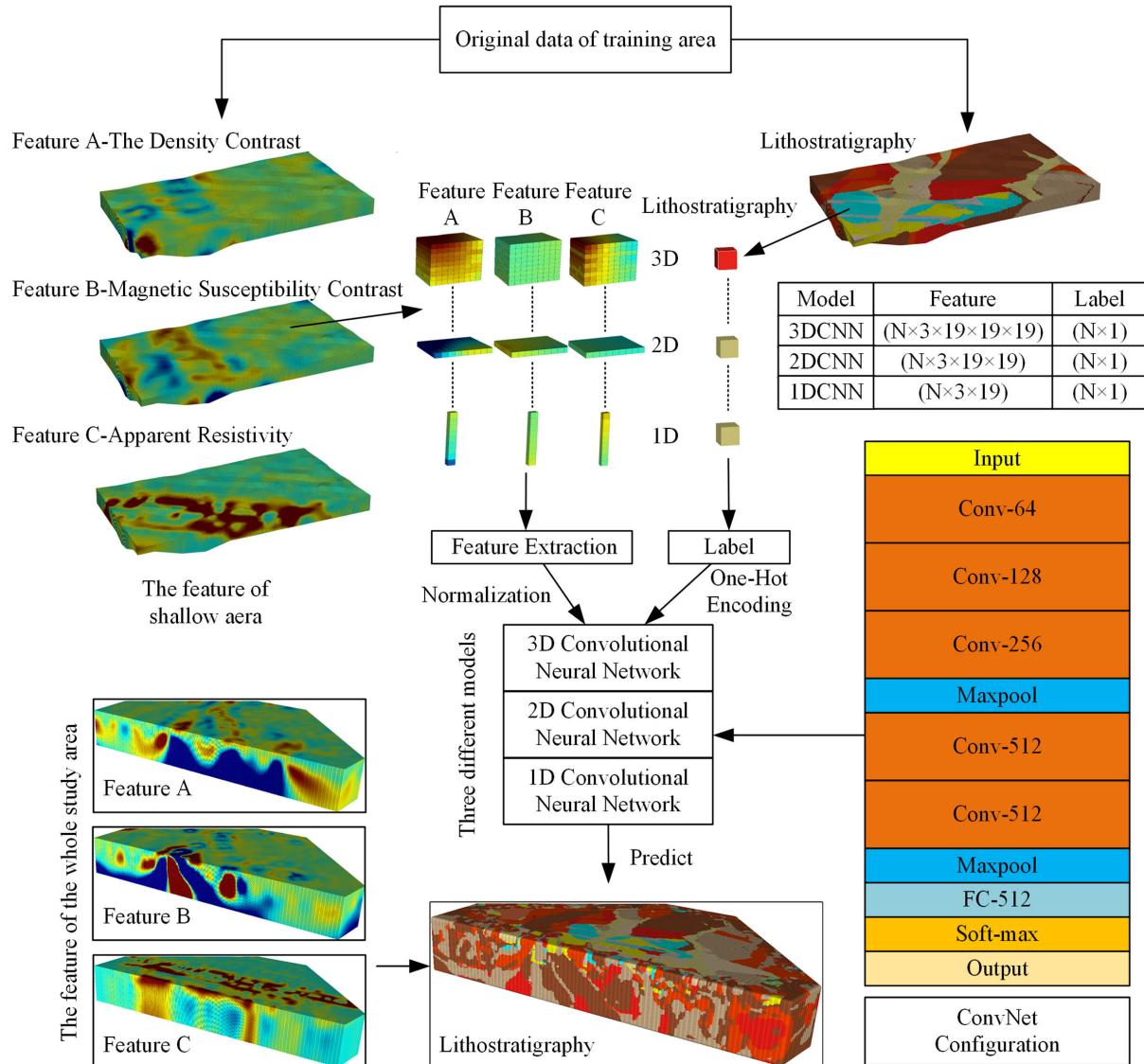


FIGURE 1: Architecture of the 1D, 2D, and 3D CNN-based pseudo-lithostratigraphic modeling.

time-consuming, and inadequate for identifying deep subsurface litho-strata in a large area. Additionally, human subjectivity and experience wield considerable influence over these conventional approaches' outcomes. Hence, there exists a need for more reliable lithostratigraphic identification methods using physical or chemical properties. Three-dimensional geological modeling plays a vital role in portraying subsurface spatial characteristics, for example, litho-strata, fault networks, physical and chemical properties, and quantitative mineralization [1–3]. With the advancements in deep-penetration geophysical and geochemical exploration, the prospect of studying geological bodies' physical and chemical properties from a 3D perspective becomes feasible [4–6]. These methodologies deepen the comprehension of the lithostratigraphy's spatial distribution patterns and guide mineral resource exploration.

In recent years, advancements in computing capacity have empowered geologists to pursue more intelligent

lithostratigraphic identification. Lithostratigraphic identification can be framed as a classification problem, and a variety of machine-learning methods have been employed for litho-stratigraphic classifications of 1D well-logging and 2D geological maps or images. Support vector machines, capable of transforming linearly indistinguishable samples into distinguishable ones, encounter practical challenges with large-scale training samples [7, 8]. Decision tree-based ensemble methods, for example, random forests [9–11] and gradient-boosting decision tree [12–14], overlook cross-correlation between different attributes in the dataset despite being less affected by category imbalance. Generally, convolutional neural networks (CNNs), as a deep learning method, can enhance the efficiency and scalability of 1D and 2D lithostratigraphic identification [15–17]. Based on neural networks, CNNs can preserve the input's neighborhood correlations and spatial localities in their latent higher-level feature representations. Meanwhile, feature extraction using CNNs enables users to achieve the feature

extraction without fully understanding the specific features. CNN techniques can significantly improve 2D lithostratigraphic identification accuracy [18]. The visual geometry group (VGG) network, a CNN with a deep network structure and small convolution kernels, excels in capturing detailed features of 2D remote-sensing images while controlling computational parameters to manage workload [19].

Three-dimensional convolution was initially proposed for learning spatiotemporal features [20], in which each convolution kernel can obtain a 3D feature. The 3D kernel is not constrained to the length and width of 2D but slides across all dimensions of the input feature map, producing a 3D output tensor instead of a 2D matrix [21, 22]. Compared with a 2D CNN, a 3D CNN can be implemented by combining 3D convolution kernels into a cube formed by stacking multiple slices, thereby capturing 3D spatial features from the 3D dataset [23]. In subsurface geological pseudo-modeling, 3D CNNs have been instrumental in detecting objects from 3D seismic images. For instance, Shi et al. [24] proposed a 3D CNN model for automatic salt body detection in a 3D single-channel seismic amplitude by analyzing learned visual features [24]. Wu et al. [25] introduced a 3D CNN model for detecting and extracting faults and horizons from 3D seismic images by analyzing structural features learned from simulated 3D structural models and synthetic seismic images [25]. However, limited research has applied 3D CNNs to other geophysical properties inverted from gravitational, magnetic, and electromagnetic explorations.

This study aims to extend VGG's capabilities to 3D pseudo-modeling for identifying lithostratigraphy in deep subsurface space, with a maximum depth of ~2000 m, using a 3D dataset featuring multichannel inverted gravitational, magnetic, and electromagnetic properties in the Qingniandian area of Heilongjiang Province, NE China. Our approach involves learning the nonlinear mapping relationship between the variations of multiple physical properties and a manually constructed lithostratigraphic model in the shallow subsurface, with a maximum depth of ~200 m. The experimental results validate the reliability of the proposed 3D CNN method in lithostratigraphic identification, enabling automatic identification of related deep subsurface lithostratigraphy in a large area.

2. Three-Dimensional CNNs for Pseudo-Lithostratigraphic Modeling

In this study, we employ three geophysical properties (density contrast, magnetic susceptibility contrast, and resistivity) to train a 3D CNN-based pseudo-lithostratigraphic classification model, capable of identifying fourteen types of rocks or strata in deep 3D subsurface. We use these three geophysical properties as the input feature channels in the training blocks, which are partitioned from the shallow subsurface using marching cubes, and we consider the lithostratigraphy of the central voxel of the block as the block's label. During the training process, the

lithostratigraphic labels play a significant role in constraining the weight coefficients of the convolutional layer.

2.1. Architecture. While a CNN can directly learn end-to-end mappings and has achieved excellent results on 2D images [26], the training of 3D CNNs for 3D dataset demands increased computation complexity compared with 2D imagery datasets [27]. To assess recognition accuracies and training efficiencies of 1D, 2D, and 3D convolution kernels, we redesign the architecture of the 1D (along vertical direction), 2D (in horizontal plane), and 3D CNNs to suit pseudo-lithostratigraphic modeling, as shown in Figure 1. The parameter counts of the three models are detailed in online supplementary Table S1. For instance, the 3D CNN model comprises five 3D convolutional layers with the number of convolutional kernels gradually increasing from 64 to 512. Increasing the architecture depth with small convolutional filters significantly improves the training and recognition performance of convolutional networks [28–30]. For 3D CNNs, we utilize the convolution layer with a receiving field size of $3 \times 3 \times 3$, which is the smallest filter that can capture direct neighborhood characteristics, and the convolution stride was set to 1. The smaller convolution kernel can achieve the same scale-receptive field as the larger convolution kernel by stacking multiple layers. We also employ three stacked $3 \times 3 \times 3$ convolution kernels to approximate the effect of a $7 \times 7 \times 7$ convolution kernel and two stacked $3 \times 3 \times 3$ convolution kernels to approximate the effect of a $5 \times 5 \times 5$ convolution kernel in the neural network structure. This reduces the number of network parameters, deepens the network, and improves network capacity. We apply batch normalization [31, 32] after each convolution layer to standardize the output characteristic into a standard normal distribution, ensuring a consistent input distribution for the next convolution layer and avoiding the vanishing gradient problem. After the calculations of the convolution layers are complete, a max-pooling layer [33] is added to replace the feature of a single voxel with the characteristic statistics of its neighboring voxels. The max-pooling cube size is set to $2 \times 2 \times 2$, and the step size for pooling is 2. The pooling layer down-samples the output characteristics of the convolution layer into the characteristic of a single voxel, allowing for a wider range of characteristics from its neighborhood. We introduce the rectified linear unit (ReLU) activation function between the convolutional layer and the pooling layer to enhance the nonlinearity of the neural network model. To prevent overfitting and improve model generalization, we employ a dropout layer that randomly deactivates neurons after the activation function, with a dropout rate set to 0.2. Additionally, we use a fully connected layer to integrate the output characteristics of the convolutional and pooling layers to make full use of the existing high-order characteristics. After each fully connected layer, we apply the ReLU activation function and a dropout layer with a dropout rate of 0.5. Finally, the output of the fully connected layer is passed into the SoftMax layer [34, 35] to obtain the classification result, providing the predicted probability for each class.

2.2. Adam Optimizer. Since Adam optimizer can handle sparse gradient and nonstationary targets, requires less memory, and is also suitable for large datasets and high-dimensional spaces [36], we use the Adam optimizer to optimize our training network. The optimizer's purpose is to minimize the loss function by updating the weight parameters, typically to minimize the difference between the predicted and actual values [37]. The network weight θ is initialized according to a Gaussian distribution and adjusted during the iterations. The loss function was increasingly used to optimize θ during continuous iterative tuning. The Adam optimizer is an optimization algorithm based on a random gradient to assign adaptive individual learning rates to each neural network parameter [36, 38, 39]. The formula for calculating the adaptive learning rate of parameters using Adam is as follows:

$$E[g^2]_t = \beta E[g^2]_{t-1} + (1 - \beta)g_t^2 \quad (1)$$

$$\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{Eg_t^2 + \epsilon}}g_t \quad (2)$$

where g_t represents the gradient on the current mini-batch, t is the current iteration number, the coefficient β is usually set to 0.9 based on experiences, θ denotes the model weight, η is the step size, and the smoothing term ϵ typically defaults to 10^{-6} .

For Adam optimizer, setting the learning rate too low results in slow training convergence due to tiny adjustments of model updates in each iteration and may cause overfitting due to its sensitivity to noises and details in the training dataset. Conversely, a high learning rate can improve training but may lead to the problem of exploding gradient, where substantial parameter updates due to rapidly increasing gradients result in unstable model training. Hence, we set an adjustable learning rate to avoid exploding gradients and expedite training. We initially set a high learning rate to accelerate the training process in the training phase. As the number of iterations increases, the learning rate decreases. The specific formula is as follows:

$$lr_{t+1} = lr_t * \frac{1}{1 + t * decay} \quad (3)$$

where lr is the learning rate and $decay$ is the decay rate.

The Adam optimizer compares the predicted output values with the actual real values and evaluates them using multiple indicators (primarily accuracy and loss). Subsequently, the Adam optimizer calculates the difference between the predicted and actual values, updates the parameters of the network, and then iteratively trains the 3D CNNs until satisfactory results are achieved.

2.3. Model Evaluation. To evaluate the model's performance, we consider accuracy, the categorical cross-entropy loss function, and the confusion matrix. We plotted accuracy and categorial cross-entropy loss curves for the

training and testing dataset, utilizing true lithostratigraphic labels extracted from the shallow 3D geological model to ensure that the models do not overfit. Accuracy, representing overall precision, is calculated as the ratio of correctly predicted sampling voxels to the total number of sampling voxels:

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

where TP, TN, FP, and FN denote the numbers of true-positive, true-negative, false-positive, and false-negative sampling voxels in the training or testing datasets, respectively. A higher accuracy value indicates better model performance.

The categorical cross-entropy is a loss function for multiple categorical probabilities [40, 41], where the loss equals zero only when the actual label and the predicted label match; otherwise, it is a positive value. The formula for calculating categorical cross-entropy is as follows:

$$\text{loss} = -\frac{1}{N} \sum_{i=1}^N y_i \ln P_i - \sum_{i=1}^N (1 - y_i) \ln(1 - P_i) \quad (5)$$

where N represents the number of samples, y_i is a binary indicator function detecting whether the i th training pattern belongs to the true category, and P_i represents the probability that the sample is predicted to be its true label by the SoftMax layer.

A confusion matrix, also known as an error matrix, is a standard format for classification precision. The value of each element in the confusion matrix is precision, calculated as follows:

$$C_{i,j} = \frac{P_{i,j}}{Class_i} \times 100\% \quad (6)$$

where $Class_i$ represents the total number of the i th class corresponding to each row of the confusion matrix, and $P_{i,j}$ represents the number of samples belonging to the i th class that are predicted as the j th class. Therefore, the on-diagonal element $C_{i,i}$ of the confusion matrix represents the proportion of samples belonging to the i th class that are correctly predicted, while the off-diagonal element $C_{i,j}(i \neq j)$ represents the proportion of samples belonging to the i th class that are incorrectly predicted as the j th class.

3. Study Area and Datasets

3.1. Study Area. The Qingniandian area, situated in the southwestern part of Duobaoshan–Daxintun metallogenic belt, is one of the significant gold and polymetallic metallogenic belts in Heilongjiang Province, China [42–44] (Figure 2). The structural complexity in this area and its adjacent regions began to develop during the Mesozoic period. The primary structural features include NNE-trending faults followed by NW-trending faults, which exert significant structural influence on the Mesozoic-related

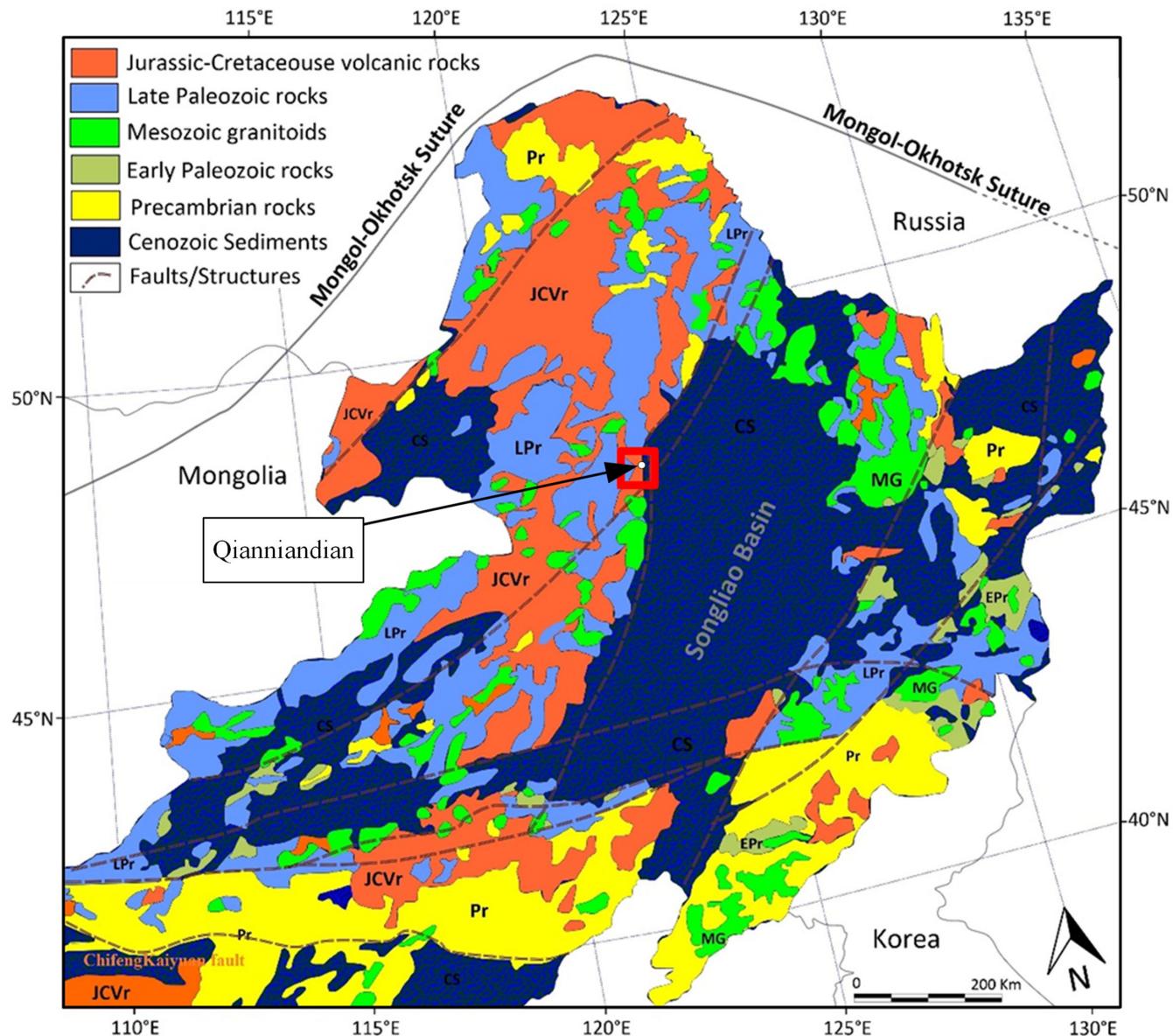


FIGURE 2: Regional geological map of NE China, modified from Reference 53.

stratigraphic units. NNE-trending faults generally exhibit straight and have zigzag structural features, serving as the primary ore-controlling structures in this area. The main period of fault formation occurred during the Mesozoic period. Hydrothermal fluids, associated with Early Cretaceous volcanic rocks, predominantly migrated along these faults. Moreover, geochemical anomalies and mineralization alterations are mainly distributed along these fault zones.

This metallogenic belt encompasses medium- to large-sized representative deposits, including the Zhengguang gold deposit, Handa gold deposit, Wudaogou gold deposit, and Tongshan copper deposit. These deposits exhibit variations in types, ranging from gold-bearing quartz vein type to fractured and altered rock type [45–47]. Their metallogenesis is also diverse, including

magmatic-hydrothermal, volcanic, and subvolcanic-hydrothermal deposits.

The lithostratigraphy of the Qingniandian area and adjacent regions mainly comprises Early Carboniferous granodiorite ($\gamma\delta C_1$), Early Carboniferous monzogranite ($\eta\gamma C_1$), Early Carboniferous syenogranite ($\xi\gamma C_1$), Lower Carboniferous Keluo complex (C_1kl), Late Carboniferous alkali feldspar granite (kyC_2), Upper Carboniferous to Lower Permian Baoligaomiao Formation (C_2P_1bl), Lower Cretaceous Jiufengshan Formation (K_1j), Lower Cretaceous Ganhe Formation (K_1g), Lower Cretaceous Gushanzhen Formation (K_1gs), Holocene Quaternary Harbin Formation (Qp^3h), Holocene Quaternary valley sediments (Qh_2), and a very small area of quartz vein (q), diorite porphyrite ($\delta\mu$), and Andesite dike (σK_1), as shown in Figure 3. These

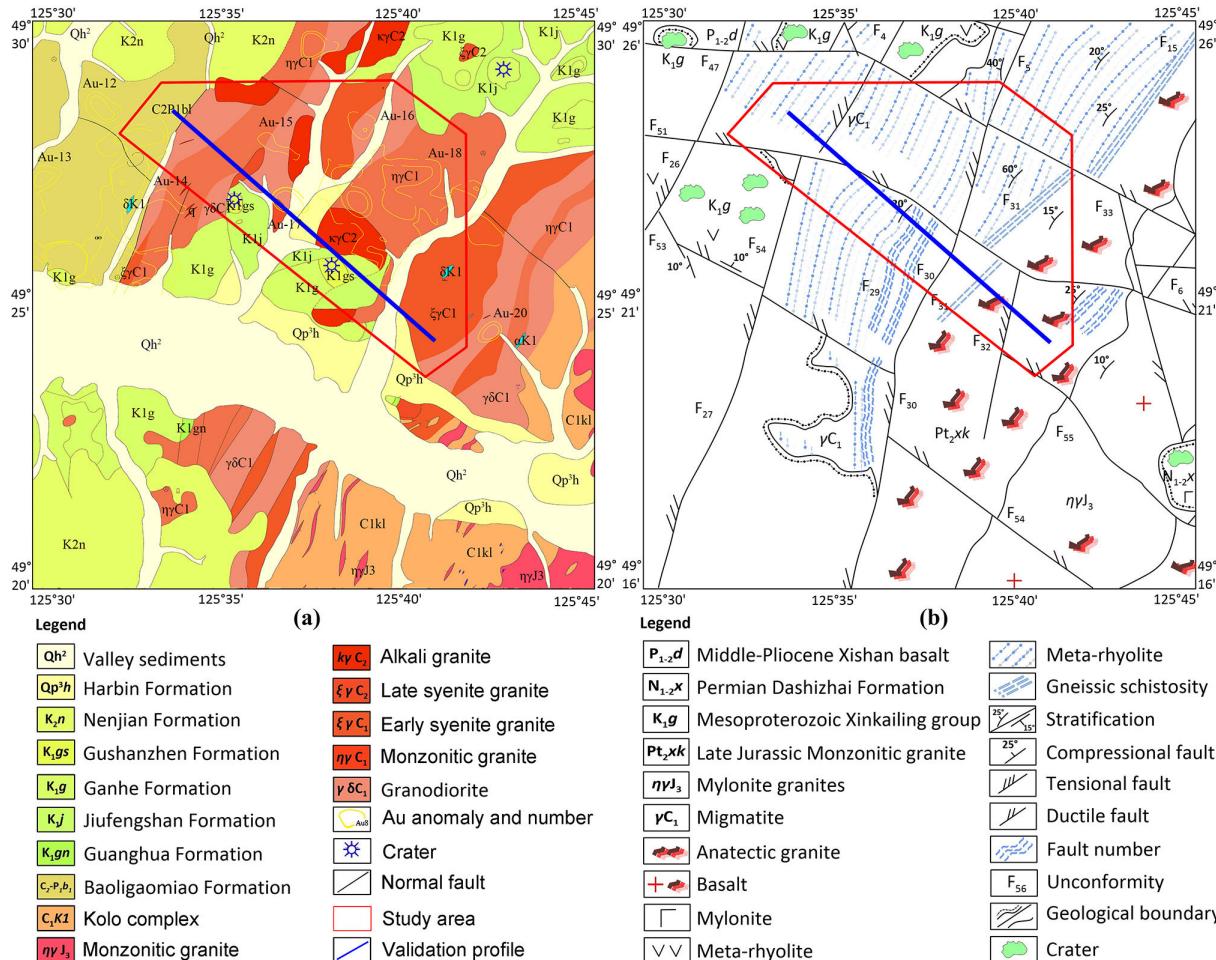


FIGURE 3: Schematic (a) geological and (b) structural maps of the Qingniandian and adjacent regions in Heilongjiang Province, NE China.

various litho-strata exhibit distinct petrographic compositions, as shown in Table 1, leading to different geophysical features.

The intrusive rocks in the study area predominantly consist of Early Carboniferous granite complex rocks and dynamic metamorphic deep intrusive rocks that typically manifest as strip and eyeball structures. The granite complex rocks are mostly mylonitic granite and granitic mylonite, with mylonitic foliation direction oriented in the northeast direction. Silicification, pyrite, and chalcopyrite are observed in granitic mylonites, which bear indicative significance for gold and copper mineral explorations. The Early Cretaceous period in the study area witnessed strong volcanic eruptive activities, primarily characterized by basic and intermediate basic volcanic activities during the Ganhe Period. These volcanic rocks were basaltic and andesitic in composition, and some underwent alteration processes involving clayization or ferritization. The hydrothermal fluid emanating from volcanic magma served as the primary thermal fluid responsible for the formation of low-temperature minerals, bearing indicative significance for polymetallic mineral explorations. Hence, recognizing the spatial distribution of these litho-strata

based on the geophysical explorations in the study area is imperative.

3.2. Datasets

3.2.1. Shallow Subsurface 3D Geological Model. To reconstruct a geological model of the shallow subsurface, ranging from ground surface (approximately 300 m) to -200 m in the study area, we initially obtained 3D fault and lithostratigraphy boundaries from the planar geological map's map-cut cross-sections. Subsequently, we tiled lithostratigraphy boundary surfaces into surfaces using triangular facets. After extending and refining the lithostratigraphy boundary surfaces, we obtained a topologically consistent geological solid model, ensuring there are no gaps or intersections between different boundary surfaces. Finally, the geological voxel model (Figure 4) was generated through rasterization from the solid model.

3.2.2. Inversed Geophysical Properties. In this study, we employed three types of inversed geophysical properties,

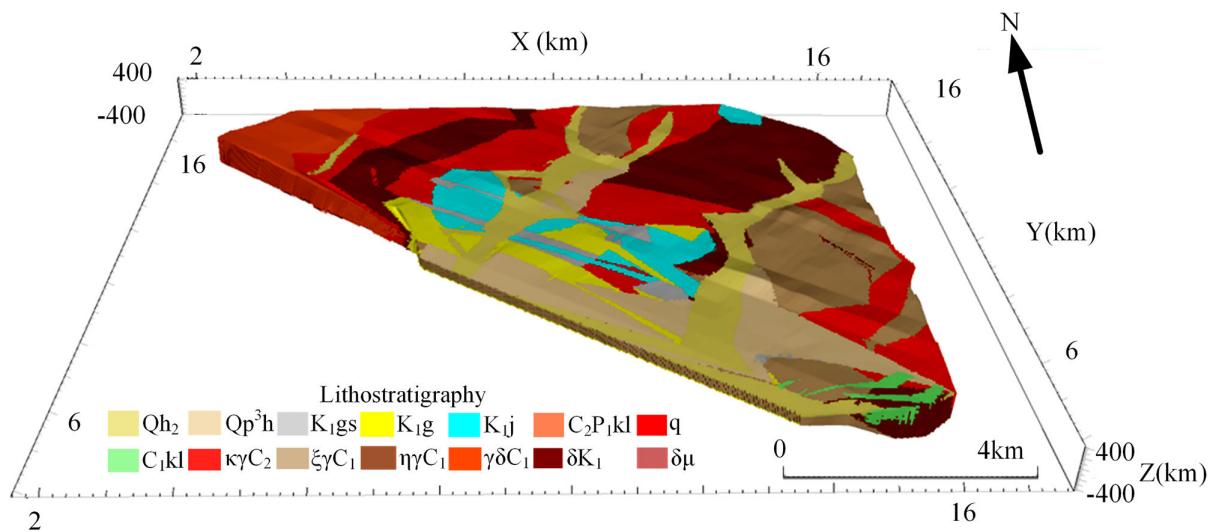


FIGURE 4: The 3D lithostratigraphy model of the shallow subsurface in the study area.

TABLE 1: Various litho-strata and their corresponding petrographic composition in the study area.

Code	Lithostratigraphy	Petrographic composition
γδC ₁	Early Carboniferous granodiorite	Plagioclase, alkaline feldspar, silica, and quartz
ηγC ₁	Early Carboniferous monzogranite	Plagioclase, potassium feldspar, quartz, hornblende, biotite, and iron titanium oxide
ξγC ₁	Early Carboniferous syenogranite	Quartz, potassium feldspar, plagioclase, and a small amount of biotite or muscovite
C ₁ kl	Lower Carboniferous Keluo complex	Biotite plagiogneiss, leptynite, granulite, amphibolite, and migmatite
κγC ₂	Late Carboniferous alkali feldspar granite	Biotite, potassium feldspar, plagioclase, and quartz
C ₂ P ₁ bl	Upper Carboniferous to Lower Permian Baoligao miao formation	Metamorphic rhyolite, meta-rhyolitic welded tuff, and light metamorphic dacite lithic-crystal tuff
K ₁ j	Lower Cretaceous Jiufengshan formation	Siltstone, fine sandstone, tuff stone, feldspathic debris sandstone, and tuff
K ₁ g	Lower Cretaceous Ganhe formation	Basalt, vesicle Andesite basalt, andesite-basalt, basaltic andesite, and trachyandesite
K ₁ gs	Lower Cretaceous Gushan formation	Rhyolite, rhyolitic tuff, and lithophysa rhyolite
Qp ³ h	Holocene Quaternary Harbin formation	Mild clay, ferromanganese bearing nodules, sand, and sandy gravel
Qh ₂	Holocene Quaternary Valley sediments	Sand, clay, and gravel
q	Quartz vein	Quartzite
δμ	Diorite porphyrite	Plagioclase and amphibole granite
δK ₁	Andesite dike	Plagioclase and hypersthene

that is, density contrast, magnetic susceptibility contrast, and resistivity, as predictors for lithostratigraphy. The 3D geophysical properties, inversed from geophysical gravitational, magnetic, and electromagnetic explorations by the Shenyang Center of China Geological Survey, serve as crucial features for lithostratigraphy recognition and ore-forming process analysis in the deep subsurface. Distinct litho-strata exhibit variations in the signals obtained from various geophysical explorations, which are reflected in the reversed geophysical features. As depicted in online supplementary Figures S1–S3, the

geophysical properties of different litho-strata, as labeled by the shallow subsurface 3D geological model of the study area, fall within distinct ranges. Variations in density, magnetism, and resistivity are attributed to the different mineral compositions of litho-strata, impacting geophysical explorations. Therefore, we explored the nonlinear mapping relationship between inversed geophysical features and lithostratigraphy to enable lithostratigraphy recognition in deep subsurface.

Although the spatial distribution of lithostratigraphy is primarily known from the 3D geological model resulting

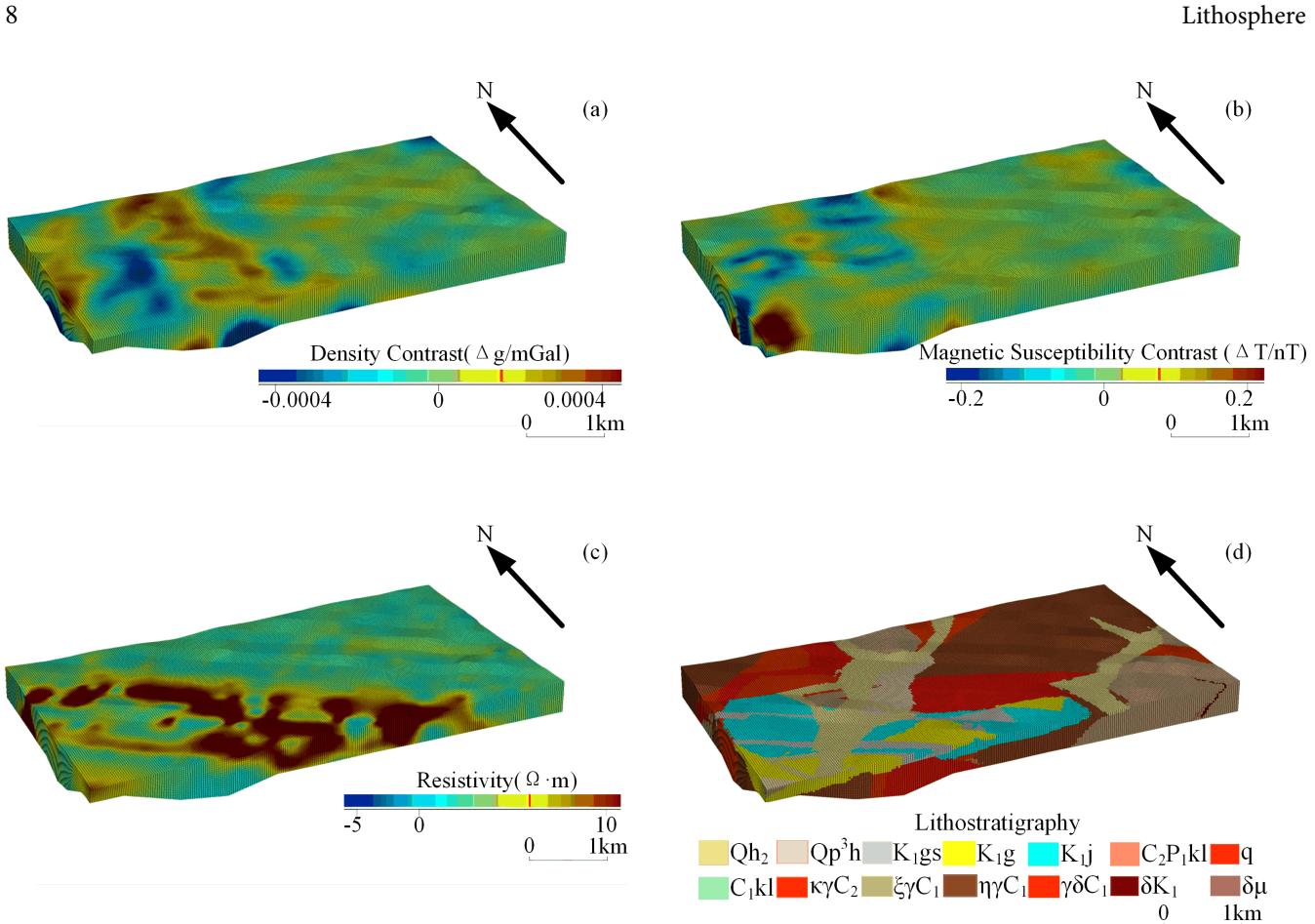


FIGURE 5: Inversed geophysical properties and lithostratigraphy distribution in the study area: (a) density contrast; (b) magnetic susceptibility contrast; (c) resistivity; and (e) lithostratigraphy.

from cross-sections in the shallow subsurface of the study area, we employed the 3D shallow subsurface geological model and inversed geophysical model as training and testing datasets, as shown in Figure 5. This involves utilizing the inversed geophysical density, magnetism, and resistivity as input features and the lithostratigraphy from the geological model as a label. The model was subsequently utilized to predict the distribution of lithostratigraphy in the deep subsurface of the study area.

Each voxel in the 3D grid model contains seven attributes, that is, central coordinates (x , y , and z), density contrast, magnetic susceptibility contrast, resistivity, and lithostratigraphic label. We chose a 3D grid measuring $570 \times 689 \times 33$ voxels from the shallow subsurface of the study area, with each voxel measuring $35 \times 35 \times 35$ m. These voxels were divided into training (9,720,068 voxels) and testing (3,240,022 voxels) datasets, maintaining a 3:1 ratio. Employing a marching cube algorithm, we partitioned the 3D grid into a series of 1D (along the Z-axis), 2D (in the XY plane), and 3D blocks, serving as the samples, with a clipping stride of 1 voxel. This strategy allows us to obtain a larger number of samples. The lithostratigraphy, labeled using the shallow subsurface 3D geological model, for the central voxel of each block created by marching cubes serves as the label for the entire block (Figure 6).

We utilized one-hot encoding to convert the lithostratigraphy label into a binary encoding format, facilitating processing by machine learning algorithms.

4. Results

To assess the models' generalization, we divided the available samples into training and testing datasets, with a ratio of 3:1 between them. The training process involves three stages: model training using the training dataset, evaluating model performance using the testing dataset, and using the trained model to predict lithostratigraphy in the deep subsurface of the study area.

We experimented with different optimizers, marching cube sizes, and training epochs, resulting in the design of seven distinct 3D CNN-based models for lithostratigraphy recognition. Through a comprehensive evaluation encompassing overall accuracy, classification precision, loss function, and prediction results, we identified an optimal 3D CNN-based model with a specific parameter combination.

4.1. Optimizer. The choice of optimizer significantly influences training outcomes. We tested three different

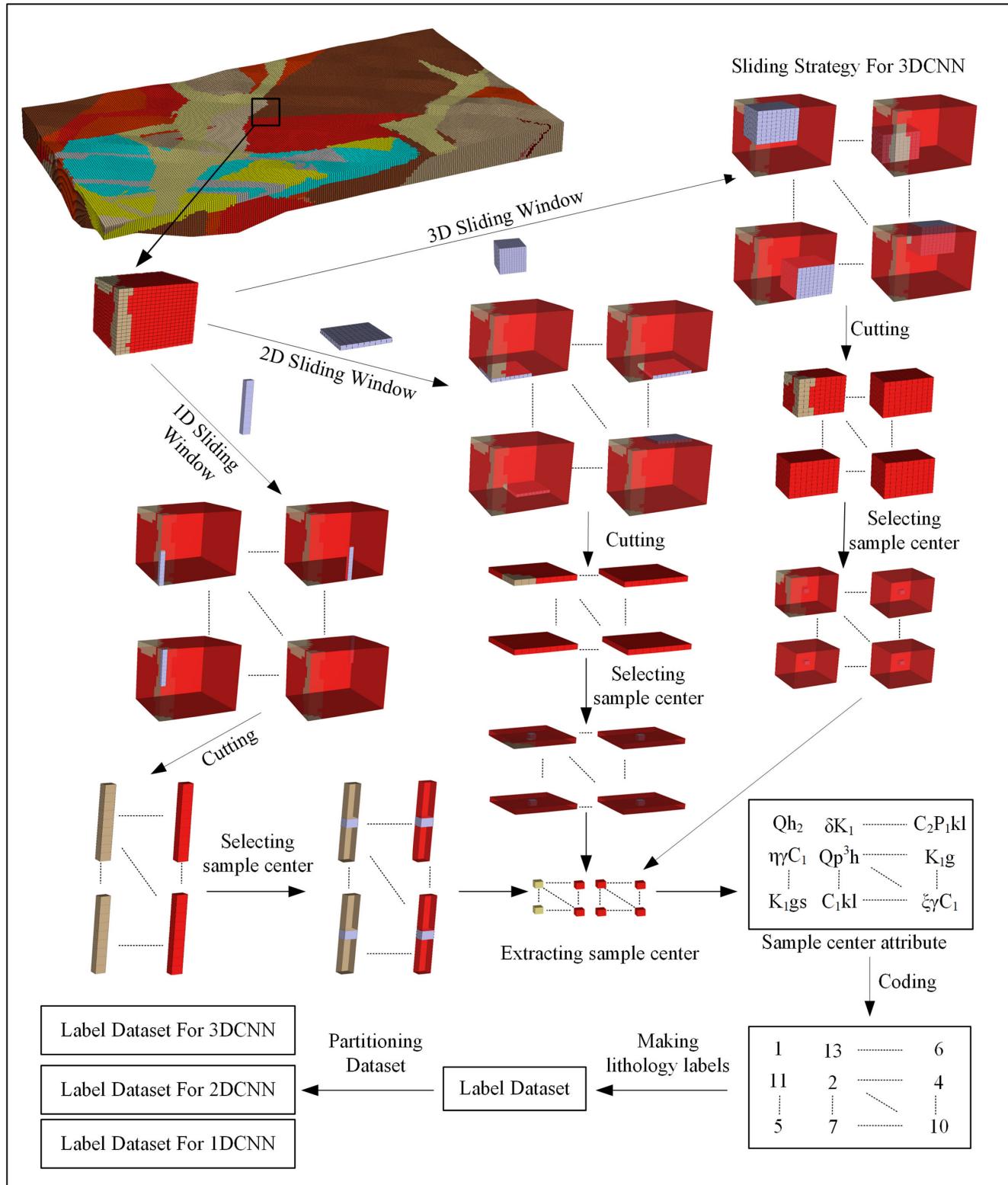


FIGURE 6: Process of creating lithostratigraphy labels for the 1D, 2D, and 3D samples.

optimizers, that is, stochastic gradient descent (SGD) [48, 49], AdaGrad [50, 51], and Adam, aiming to expedite training and enhance accuracy. As indicated in Table 2, the model employing the Adam optimizer demonstrates higher accuracy and lower loss values in

both training and testing datasets. Figures 7 and 8 illustrate that the Adam-based model converges quickly and achieves superior training results. Notably, the lower performance on the training set than that on the testing set arises from our use of the dropout option

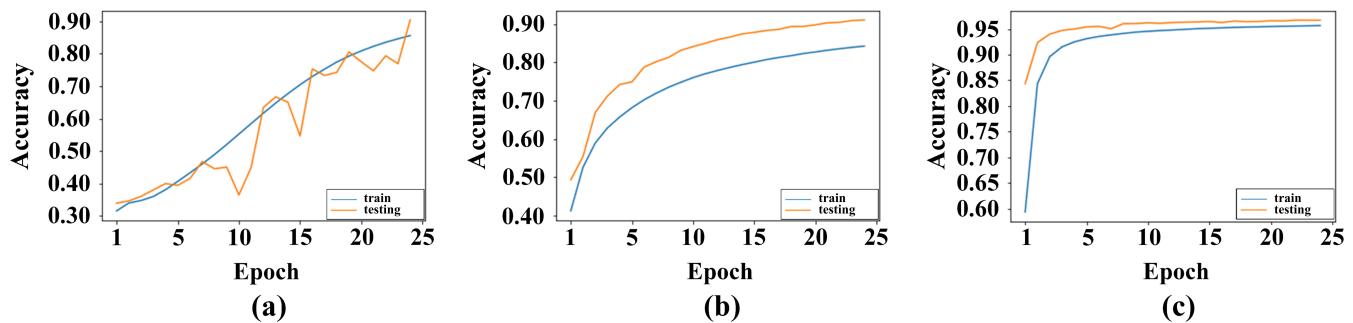


FIGURE 7: Accuracy curves obtained by the models with different optimizers: (a) SGD, (b) Adagrad, and (c) Adam.

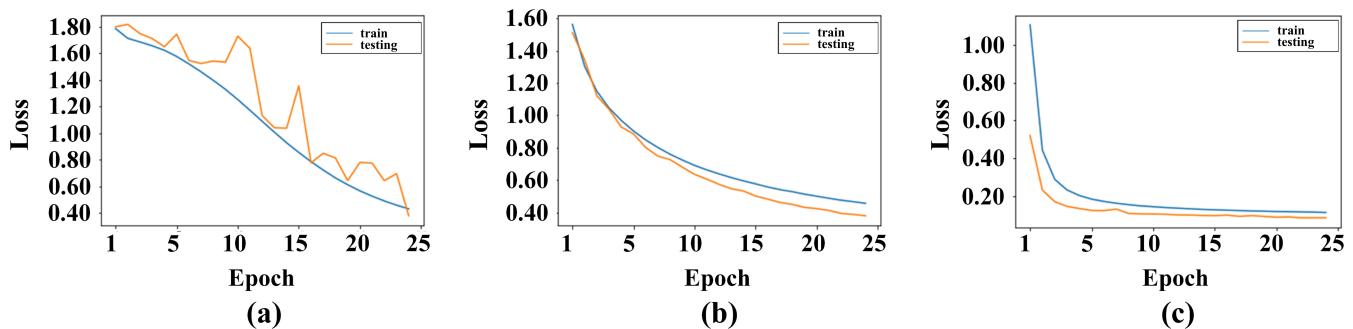


FIGURE 8: Loss curves obtained by the models with different optimizers: (a) SGD, (b) AdaGrad, and (c) Adam.

TABLE 2: Accuracy and loss values obtained by models using different optimizers.

Optimizer	Accuracy of the training dataset	Accuracy of the testing dataset	Loss of the training dataset	Loss of the testing dataset	Time consuming (hour)
SGD	85.54%	90.35%	0.4306	0.3788	6.25
AdaGrad	84.02%	91.08%	0.4559	0.3789	6.92
Adam	95.66%	96.70%	0.1120	0.0833	8.01

Hyperparameter: Dropout on; Epoch = 25; learning rate = 0.0001; Batch_size = 64; Cube size = $19 \times 19 \times 19$.

during training. The confusion matrices presented in online supplementary Figure S4 indicate that the SGD- and AdaGrad-based models struggle to accurately identify certain litho-strata, for example, quartz vein and Andesite dike in the testing dataset, which have limited training samples. However, the Adam-based model excels in identifying these litho-strata.

We compared the predicted lithostratigraphy results with the actual lithostratigraphy in the training area, as depicted in online supplementary Figure S5. The SGD- and AdaGrad-based models differ significantly from the actual distribution, whereas the Adam-based model not only reconstructs the actual distribution but also defines clear lithostratigraphic boundaries. From the predicted results in the deep subsurface, as shown in online supplementary Figure S6, the SGD- and AdaGrad-based models wrongly identify newly deposited Cretaceous lithostratigraphy in the deep subsurface, which contradicts geological laws. In contrast, the Adam-based model achieves better results in lithostratig-

raphy prediction in the deep subsurface of the study area.

4.2. Marching Cube Size. Using a larger marching cube size on the samples provides more neighbors correlated to the central voxel of the block. Consequently, we tested three different marching cube sizes, that is, $17 \times 17 \times 17$, $19 \times 19 \times 19$, and $21 \times 21 \times 21$, to optimize performance. As depicted in Table 3, increasing the cube size from $17 \times 17 \times 17$ to $19 \times 19 \times 19$ improves the model's accuracy and reduces the loss on both the training and testing datasets. This enhancement is attributed to the larger marching cube size capturing more neighborhood characteristics and subsequently optimizing the model's performance. However, when the cube size is increased from $19 \times 19 \times 19$ to $21 \times 21 \times 21$, the model's performance shows only marginal improvement. The testing curve also indicates that increasing the marching cube size expedites model convergence during training, as shown in Figures 9 and 10.

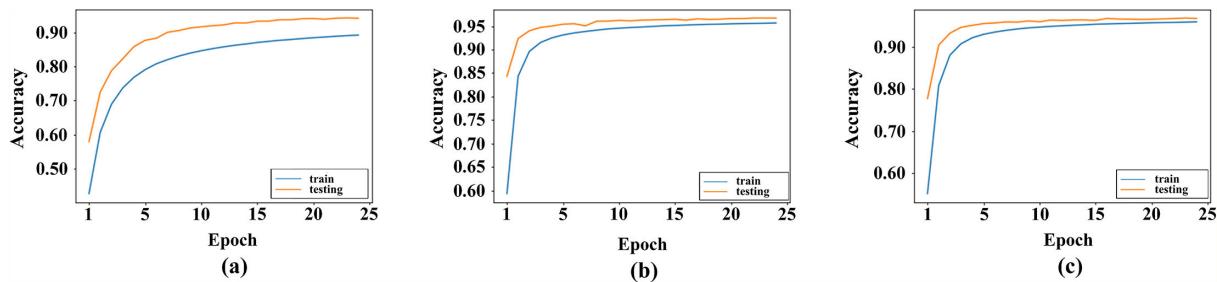


FIGURE 9: Accuracy curves of the models with different marching cube sizes: (a) $17 \times 17 \times 17$, (b) $19 \times 19 \times 19$, and (c) $21 \times 21 \times 21$.

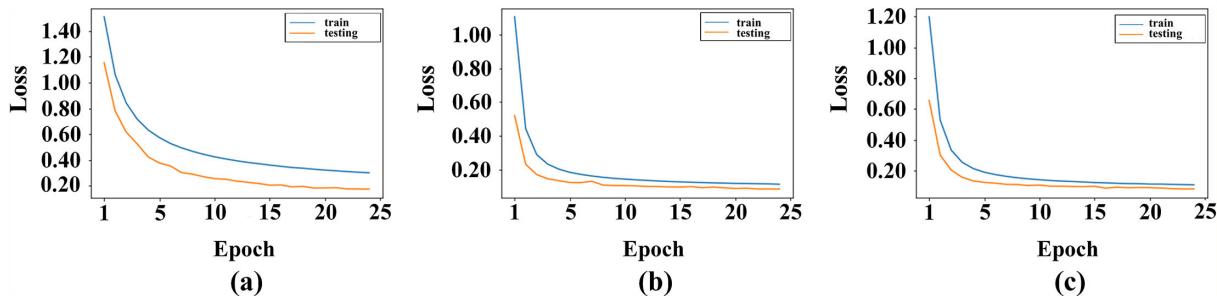


FIGURE 10: Loss curves of the models with different marching cube sizes: (a) $17 \times 17 \times 17$, (b) $19 \times 19 \times 19$, and (c) $21 \times 21 \times 21$.

TABLE 3: Accuracies and losses of the models with different marching cube sizes.

Cube size	Accuracy of the training dataset	Accuracy of the testing dataset	Loss of the training dataset	Loss of the testing dataset	Time consuming (hour)
$17 \times 17 \times 17$	89.19%	94.12%	0.3014	0.1752	7.72
$19 \times 19 \times 19$	95.66%	96.70%	0.1120	0.0833	8.01
$21 \times 21 \times 21$	95.90%	96.71%	0.1072	0.0825	8.22

Hyperparameter: Dropout on; Optimizer = Adam; Epoch = 25; learning rate = 0.0001; Batch_size = 64.

TABLE 4: Accuracies and losses of models with different numbers of epoch.

Training epochs	Accuracy of the training dataset	Accuracy of the testing dataset	Loss of the training dataset	Loss of the testing dataset	Time consuming (hour)
10	93.99%	95.78%	0.1072	0.1619	3.20
25	95.66%	96.70%	0.1120	0.0833	8.01
50	95.62%	96.89%	0.1236	0.0778	15.37

Hyperparameter: Dropout on; Optimizer = Adam; learning rate = 0.0001; Batch_size = 64; Cubes size = $19 \times 19 \times 19$.

The confusion matrices of the three models, as shown in online supplementary Figure S7, indicate that increasing the cube size from $17 \times 17 \times 17$ to $19 \times 19 \times 19$ significantly improved the model's accuracy. While the model with a marching cube size of $21 \times 21 \times 21$ displays superior performance metrics, its predicted complex lithological boundaries lack sufficient rationality. Nonetheless, increasing the cube size from $19 \times 19 \times 19$ to $21 \times 21 \times 21$ does not yield a noticeable improvement in model accuracy. Compared with the model with a marching cube size of $17 \times 17 \times 17$, the prediction result of the model with a marching cube size of $19 \times 19 \times 19$ exhibits better consistency with the actual lithostratigraphy (online

supplementary Figure S8). The models with marching cube sizes of $17 \times 17 \times 17$ and $21 \times 21 \times 21$ incorrectly identify artifacts where several Cretaceous and Quaternary litho-strata appear in the deep subsurface of the study area (online supplementary Figure S9). Hence, the optimal result was obtained with the model with a marching cube size of $19 \times 19 \times 19$.

4.3. Training Epochs. We designed three models with training epochs of 10, 25, and 50. Increasing the training epochs from 10 to 25 gradually improved model performance (Table 4, Figures 11, 12, and S10). However, beyond twenty-five epochs, further increases do not significantly

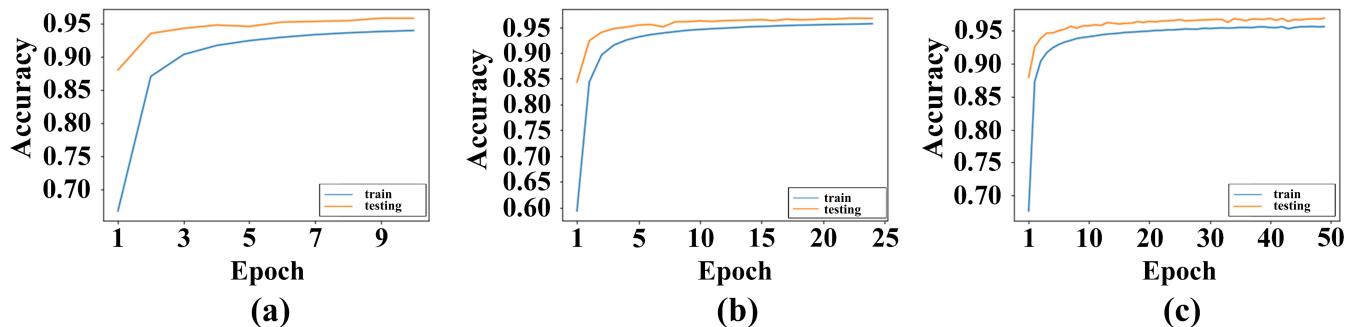


FIGURE 11: Accuracy curves of the models with different training epochs: (a) 10, (b) 25, and (c) 50.

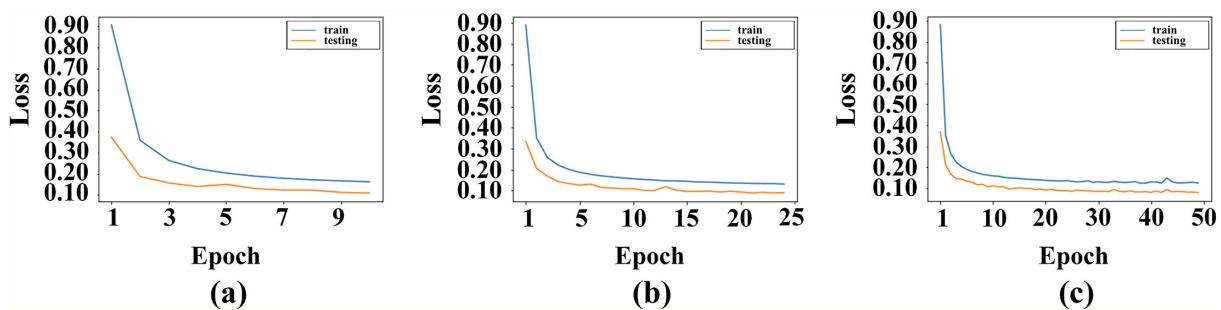


FIGURE 12: Loss curves of the models with different training epochs: (a) 10, (b) 25, and (c) 50.

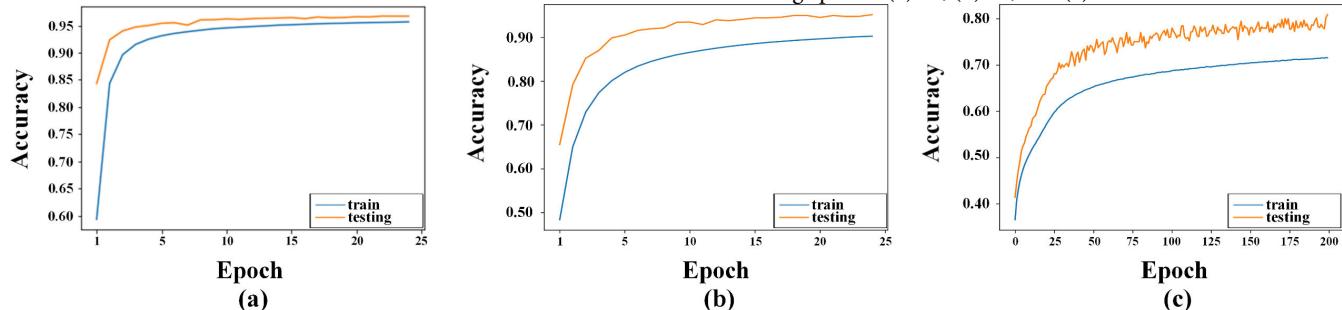


FIGURE 13: Accuracy curves of the (a) 3D, (b) 2D, and (c) 1D CNN models.

TABLE 5: Accuracies and losses of models with different CNN structures.

CNN structure	Accuracy of the training dataset	Accuracy of the testing dataset	Loss of the training dataset	Loss of the testing dataset	Time consuming (hour)
3D CNN	95.66%	96.70%	0.1120	0.0833	8.01
2D CNN	90.25%	95.17%	0.2659	0.1396	5.82
1D CNN	71.55%	80.93%	0.7543	0.5269	40.56

3D CNN: Dropout on; Optimizer = Adam; learning rate = 0.0001; Batch_size = 64; cubes size = $19 \times 19 \times 19$; Epoch = 25.

2D CNN: Dropout on; Optimizer = Adam; learning rate = 0.0002; Batch_size = 64; cubes size = 19×19 ; Epoch = 25.

1D CNN: Dropout on; Optimizer = Adam; learning rate = 0.0001; Batch_size = 64; cubes size = 1×19 ; Epoch = 200.

enhance model performance, as the model has already converged.

As training epochs increased from 10 to 25, the model's lithostratigraphy prediction results increasingly aligned with the actual lithostratigraphy distribution (online supplementary Figures S11 and S12). Nevertheless, when the training epochs reach fifty, the prediction results in the deep subsurface appear inconsistent with

geological principles. Furthermore, an increased number of training epochs inevitably leads to longer model training times.

5. Discussion

5.1. Comparison of 1D, 2D, and 3D CNN Models. The convolutional layers play a pivotal role in extracting

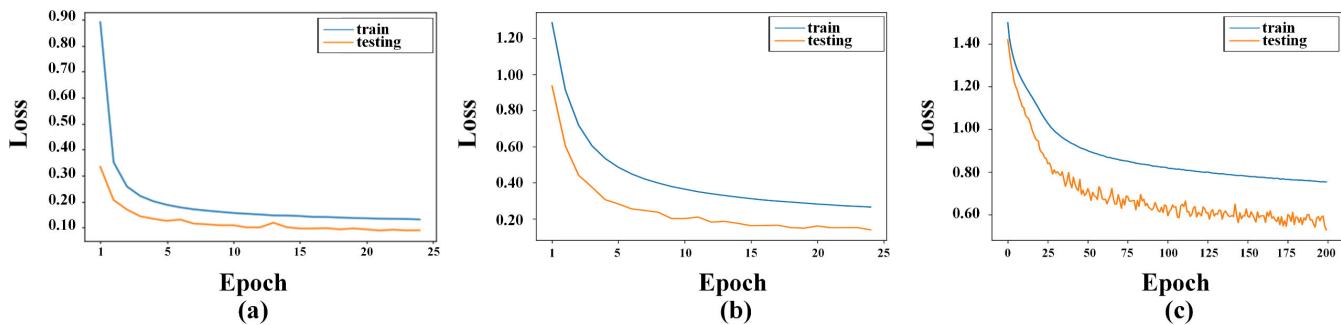


FIGURE 14: Loss curves of the (a) 3D, (b) 2D, and (c) 1D CNN models.

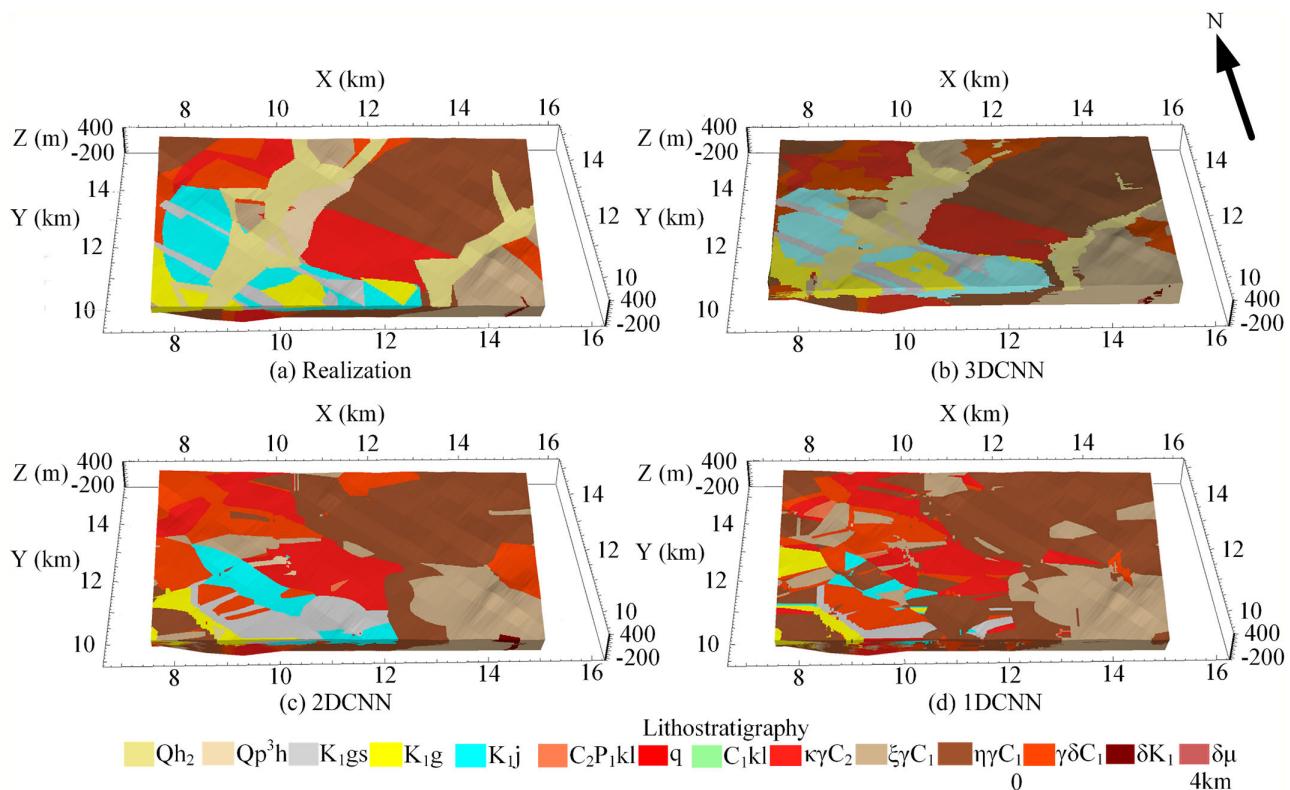


FIGURE 15: Comparison of (a) the actual lithostratigraphy distribution with predicted lithostratigraphy results from the (b) 3D, (c) 2D, and (d) 1D CNN models.

characteristics and learning the mapping correlation between geophysical properties and lithostratigraphic model. When the CNNs process input data, they capture spatial correlation characteristics within the study area, creating a large receptive field. By employing 3D convolution kernels on the samples, a larger set of correlated neighbors around the central voxel of the block is utilized compared with 2D and 1D convolution kernels. We evaluated the performance of three distinct convolution kernels, that is, 1D, 2D, and 3D. As depicted in Table 5, the 3D CNN model exhibits the highest accuracy and lowest loss values across both training and testing datasets. This best performance stems from the 3D convolution kernel's capacity to capture a more comprehensive set of neighborhood characteristics. As illustrated in Figures 13 and 14, after twenty-five epochs, further iterations

do not significantly enhance the performance of the 2D and 3D CNN models as they have already converged. In contrast, the 1D CNN model converges slowly, requiring 200 epochs, and achieves the lowest performance. Moreover, an increased number of training epochs leads to longer training times for the 1D CNN model. The confusion matrices, as shown in online supplementary Figure 13, highlight that the 3D CNN markedly enhances the model's accuracy. Conversely, the 1D and 2D CNN models exhibit inferior performance metrics, with their predicted lithological boundaries displaying insufficient rationality. As shown in Figure 15, a comparison between the predicted lithostratigraphy results with the actual lithostratigraphy in the training area reveals significant discrepancies between the 1D and 2D CNN models. Conversely, the 3D CNN model not only reconstructs the actual distribution

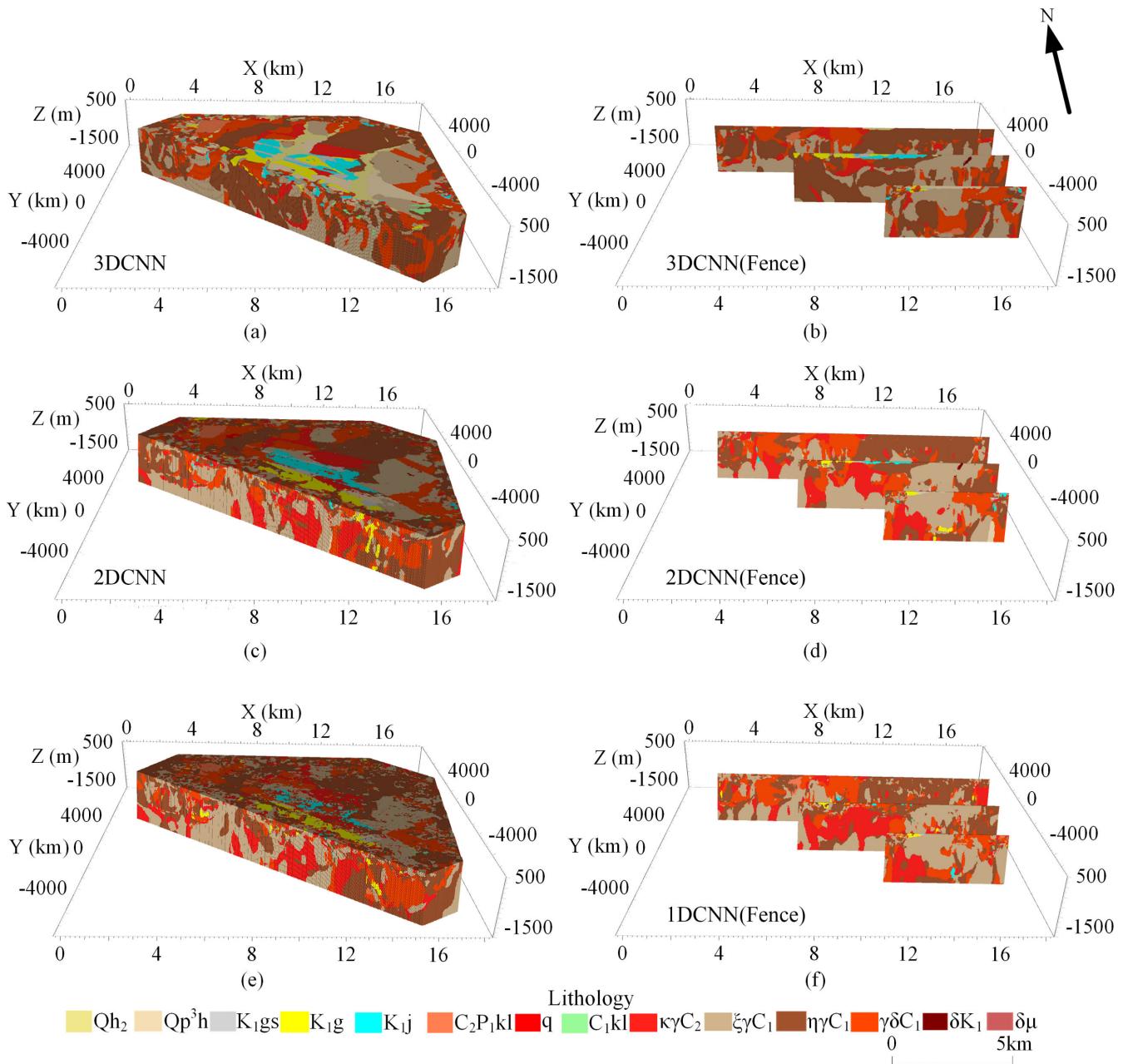


FIGURE 16: Comparison between the predicted litho-strata and corresponding cross-sections by the (a and b) 3D, (c and d) 2D, and (e and f) 1D CNN models.

but also delineates distinct lithostratigraphic boundaries, demonstrating better alignment with the actual lithostratigraphy.

Examining the predicted results in the deep subsurface, as shown in Figure 16, both the 1D and 2D CNN models inaccurately identify newly deposited Cretaceous lithostratigraphy in the deep subsurface, a contradiction to geological laws. Conversely, the optimal result in the deep subsurface is achieved by the 3D CNN model with its ability to accurately capture subsurface lithostratigraphic characteristics.

5.2. Validations of 3D CNN Model. In this study, we applied the 3D Adam-CNN model to classify fourteen types of

litho-strata. The experimental results demonstrate that the model using the Adam optimizer fits the actual lithostratigraphic conditions better and achieves convergence within a relatively short training period. The model with a marching cube size of $19 \times 19 \times 19$ is sufficiently large to capture neighborhood-inversed geophysical features for optimal performance. Furthermore, increasing the number of training epochs does not significantly improve overall accuracy and prediction precision after reaching 25.

After parameter tuning and comparisons, the model utilizing the Adam optimizer, a marching cube size of $19 \times 19 \times 19$, and twenty-five training epochs yields the best prediction results. We compared the statistical distribution difference of three geophysical properties in several

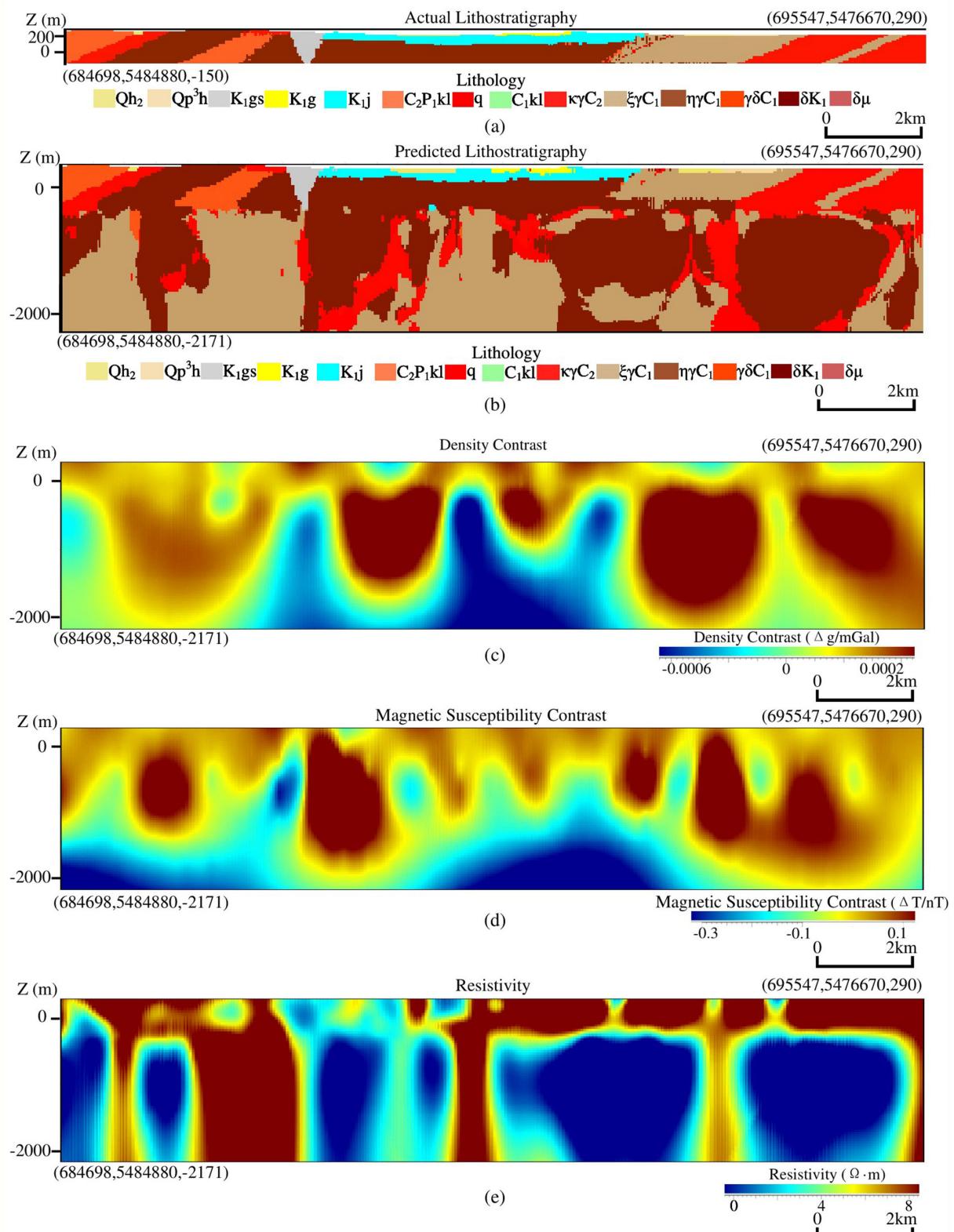


FIGURE 17: Comparison between the actual and predicted lithostratigraphic distributions on the validation profile with the corresponding inversed geophysical properties: (a) actual lithostratigraphy in the shallow subsurface; (b) predicted lithostratigraphy; (c) density contrast; (d) magnetic susceptibility contrast; and (e) resistivity.

actual and predicted litho-strata using histograms and fitted normal distribution curves (online supplementary Figure

14). The comparison results indicate that the distribution of three geophysical properties in several actual and

predicted litho-strata is roughly consistent, particularly in the Cretaceous and Carboniferous litho-strata.

We selected a validation profile in the study area and compared the real lithostratigraphic distributions in the shallow subsurface on the profile with the predicted lithostratigraphic distributions by the model and the corresponding inversed geophysical properties on the profile, as shown in Figure 17. The validation profile indicates that the predicted lithostratigraphy results are relatively reliable and closely match the actual lithostratigraphic distribution, especially in the early Carboniferous granite, which is a crucial indicator for metallogenic prediction. In the validation profile, the lithostratigraphic distributions of deep and shallow subsurface spaces in the study area exhibit good continuity, with clear and smooth boundaries between various litho-strata, conforming to the sequence of geological sedimentation. The models achieve high accuracy in predicting granite and Cretaceous strata due to their unique mineral compositions and geophysical features. However, the models have high classification error rates in quartz veins and often misidentify them as granites due to two primary reasons: (1) the inversed geophysical properties of these quartz veins are similar to those of granite, making it difficult to distinguish the two litho-strata; and (2) these quartz veins only exist in a few areas with low thickness and a small number of samples, limiting the model's ability to capture sufficient corresponding characteristics to fit the nonlinear relationship between geophysical properties and lithostratigraphy. Our 3D Adam-CNN model successfully identifies and analyzes the differences in three inversed geophysical features among different litho-strata, learning the mapping relationships between them to achieve correct lithostratigraphic modeling. Consequently, it infers the lithostratigraphic distribution in the deep subsurface of the study area. Given the lithostratigraphic distribution is closely linked to the mineral deposits, particularly certain litho-strata (e.g., $\gamma\delta C_1$, $\eta\gamma C_1$, and $\xi\gamma C_1$), considering their contact zones as prospecting indicators, the lithostratigraphic prediction results obtained through the proposed 3D Adam-CNN method can serve as essential indicators for mineral prospecting.

6. Conclusions

This study focuses on pseudo-lithostratigraphic modeling using inversed geophysical data based on the 3D CNN model, achieving satisfactory experimental results with high performance (e.g., an overall accuracy of 96.7%). The model's predictions in the study area align with the geological chronological sequence, demonstrating the proposed 3D Adam-CNN model's robust generalization ability. In conclusion, this study contributes as follows:

- (1) Our approach leverages 3D geophysical properties as input features for training while employing the shallow subsurface 3D geological model as labels, which stands as a reliable pathway for pseudo-lithostratigraphic modeling in the deep subsurface.

- (2) Compared with the 1D and 2D CNN models, our proposed 3D CNN model significantly elevates the accuracy of pseudo-lithostratigraphic modeling, providing valuable guidance for mineral prospecting in our study area.
- (3) By comparing the predicted lithostratigraphic distributions with actual lithostratigraphic distributions in the validation profile, our 3D CNN model exhibits strong agreement with the actual lithostratigraphic model in the shallow subsurface dataset, effectively discerning boundaries and trends among different litho-strata in the deep subsurface.

There are several limitations that need to be addressed. A fundamental assumption of our proposed approach is that the same set of litho-strata distributes in both shallow subsurface and deep subsurface. However, the inherent inhomogeneity and anisotropy of the Earth's subsurface weaken this generalization, particularly when training a model using a shallow subsurface dataset to predict the deep subsurface where litho-strata vary significantly. While our proposed 3D CNN framework displays applicability to lithostratigraphic modeling in other countries and regions, the specific workload required for labeling and training also diminishes its practicality. Furthermore, concerns about the interpretability of CNN approaches [52] and the need to elucidate the workings of CNNs to enhance algorithmic transparency of pseudo-lithostratigraphic modeling have been raised. Efforts to visualize the internal mechanism of CNNs and improve the understanding of the nonlinear mapping between lithostratigraphy and inversed geophysical properties are imperative. Moreover, integrating deep-level features and neighborhood correlations in the 3D CNN model has aimed to reduce subjectivity and reliance on simplistic statistical features in lithostratigraphic identification using inversed geophysical properties characterized by multiple interpretations. However, future validation should also consider borehole data and corresponding geomechanical properties to validate the predicted results and assess the model's generalizability.

Data Availability

The dataset of the current study is not publicly available due to a data privacy agreement we signed with the Shenyang Center of China Geological Survey but is available from the corresponding author upon reasonable request.

Conflicts of Interest

We confirm that there is no conflict of interest.

Acknowledgments

The authors express their gratitude to Dr. Zhang Chun-peng and Dr. Liu Bao-shan (Shenyang Center of China Geological Survey) and Dr. Liu Wen-yu (East China University of Technology) for their kind assistance with the data

collection. Special thanks are extended to the MapGIS Laboratory Co-Constructed by the National Engineering Research Center for Geographic Information System of China and Central South University for providing the MapGIS software (Wuhan Zondy Cyber-Tech Co. Ltd., Wuhan, China).

This study was supported by grants from the National Natural Science Foundation of China (Grant No. 42072326) and China Geological Survey Project (Grant No. DD20190156).

Supplementary Materials

Figures S1–S3 depict the distinct ranges of the three geophysical properties corresponding to different lithostrata. Table S1 itemizes the parameter counts for the 1D, 2D, and 3D CNN models. Figures S4–S6 present comparisons among models using different optimizers. Figures S7–S9 showcase comparisons based on varying marching cube sizes. Figures S10–S12 display model performances across different training epochs. Figure S13 demonstrates the confusion matrices generated in testing dataset for the 1D, 2D, and 3D CNN models. Figure S14 illustrates the comparison between the distribution of geophysical properties in the actual and predicted litho-strata.

References

- [1] C. Lajaunie, G. Courrioux, and L. Manuel, “Foliation fields and 3D cartography in geology: principles of a method based on potential interpolation,” *Mathematical Geology*, vol. 29, no. 4, pp. 571–584, 1997.
- [2] B. Zhang, Y. Chen, A. Huang, H. Lu, and Q. Cheng, “Geochemical field and its roles on the 3D prediction of concealed ore-bodies,” *Yanshi Xuebao/Acta Petrologica Sinica*, vol. 34, pp. 352–362, 2018.
- [3] L. Wang, X. Wu, B. Zhang, et al, “Recognition of significant surface soil Geochemical anomalies via weighted 3d shortest-distance field of subsurface Orebodies: a case study in the Hongtoushan copper mine, NE China,” *Natural Resources Research*, vol. 28, no. 3, pp. 587–607, 2019.
- [4] Q. Wu, H. Xu, and X. Zou, “An effective method for 3D geological modeling with multi-source data integration,” *Computers & Geosciences*, vol. 31, no. 1, pp. 35–43, 2005.
- [5] X. Hu, X. Li, F. Yuan, et al., “3D numerical simulation-based targeting of skarn type mineralization within the Xuan-cheng-Magushan orefield, middle-lower Yangtze metallogenetic belt, China,” *Lithosphere*, vol. 2020, no. 1, 2020.
- [6] B. Zhang, L. Du, U. Khan, Y. Tong, L. Wang, and H. Deng, “AdaHRBF v1.0: gradient-adaptive Hermite–Birkhoff radial basis function Interpolants for three-dimensional stratigraphic implicit modeling,” *Geoscientific Model Development*, vol. 16, no. 13, pp. 3651–3674, 2023.
- [7] T. Meng, R. Du, Z. Hou, J. Yang, and G. Zhao, “THz spectra-based SVM prediction model for yungang grottoes samples,” *Journal of Archaeological Science*, vol. 55, pp. 280–285, 2015.
- [8] M. Imran, S. Ahmad, A. Sattar, and A. Tariq, “Mapping sequences and mineral deposits in poorly exposed lithologies of inaccessible regions in azad Jammu and Kashmir using SVM with ASTER satellite data,” *Arabian Journal of Geosciences*, vol. 15, no. 6, p. 538, 2022.
- [9] M. J. Cracknell and A. M. Reading, “The upside of uncertainty: identification of lithology contact zones from airborne geophysics and satellite data using random forests and support vector machines,” *Geophysics*, vol. 78, no. 3, pp. WB113–WB126, 2013.
- [10] J. Sun, Q. Li, M. Chen, et al., “Optimization of models for a rapid identification of lithology while drilling - a win-win strategy based on machine learning,” *Journal of Petroleum Science and Engineering*, vol. 176, pp. 321–341, 2019.
- [11] Y. Ao, L. Zhu, S. Guo, and Z. Yang, “Probabilistic logging lithology characterization with random forest probability estimation,” *Computers & Geosciences*, vol. 144, p. 104556, 2020.
- [12] B. Zhang, M. Li, W. Li, et al., “Machine learning strategies for lithostratigraphic classification based on geochemical sampling data: a case study in area of Chahanwusu river, Qinghai province, China,” *Journal of Central South University*, vol. 28, no. 5, pp. 1422–1447, 2021.
- [13] Y. Zou, Y. Chen, and H. Deng, “Gradient boosting decision tree for lithology identification with well logs: a case study of Hazoxian gold deposit, Shandong peninsula, China,” *Natural Resources Research*, vol. 30, no. 5, pp. 3197–3217, 2021.
- [14] Y. Huan, L. Song, U. Khan, and B. Zhang, “Stacking ensemble of machine learning methods for landslide susceptibility mapping in Zhangjiajie city, Hunan province, China,” *Environmental Earth Sciences*, vol. 82, no. 1, p. 35, 2023.
- [15] Y. Liu and L. J. Durlofsky, “3D CNN-PCA: a deep-learning-based parameterization for complex geomodels,” *Computers & Geosciences*, vol. 148, p. 104676, 2021.
- [16] S. Shi, M. Li, W. Gao, G. Shi, J. Bai, and J. Zuo, “Improved Unet in lithology identification of coal measure Strata,” *Lithosphere*, vol. 2022, no. Special 12, 2022.
- [17] B. Zhang, M. Li, Y. Huan, U. Khan, L. Wang, and F. Wang, “Bedrock mapping based on terrain weighted directed graph convolutional network using stream sediment geochemical samplings,” *Transactions of Nonferrous Metals Society of China*, vol. 33, no. 9, pp. 2799–2814, 2023.
- [18] S. W. A. Canchumuni, J. D. B. Castro, J. Potratz, A. A. Emerick, and M. A. C. Pacheco, “Recent developments combining ensemble smoother and deep generative networks for facies history matching,” *Computational Geosciences*, vol. 25, no. 1, pp. 433–466, 2021.
- [19] S. Liu and W. Deng, “Very deep convolutional neural network based image classification using small training sample size,” in *2015 3RD IAPR Asian conference on pattern recognition (ACPR)*, 2015.
- [20] D. Tran, L. Bourdev, R. Fergus, L. Torresani, and M. Paluri, “Learning spatiotemporal features with 3D convolutional networks,” in *2015 IEEE international conference on computer vision (ICCV)*, 2015.
- [21] M. Rogozinski, J. A. C. Martinez, and R. Q. Feitosa, “3D convolution for multiday crop recognition from multitemporal image sequences,” *International Journal of Remote Sensing*, vol. 43, nos. 15–16, pp. 6056–6077, 2022.
- [22] B. Zhang, K. Xu, U. Khan, X. Li, L. Du, and Z. Xu, “A lightweight convolutional neural network with end-to-end learning for three-dimensional mineral prospectivity

- modeling: A case study of the Sanhetun Area, Heilongjiang Province, Northeastern China," *Ore Geology Reviews*, vol. 163, p. 105788, 2023.
- [23] Q. Huo, R. Song, D. Lei, et al., "Demonstration of 3D convolution kernel function based on 8-layer 3D vertical resistive random access memory," *IEEE Electron Device Letters*, vol. 41, no. 3, pp. 497–500, 2020.
- [24] Y. Shi, X. Wu, and S. Fomel, "SaltSeg: automatic 3D salt segmentation using a deep convolutional neural network," *Interpretation*, vol. 7, no. 3, pp. SE113–SE122, 2019.
- [25] X. Wu, Z. Geng, Y. Shi, N. Pham, S. Fomel, and G. Caumon, "Building realistic structure models to train convolutional neural networks for seismic structural interpretation," *Geophysics*, vol. 85, no. 4, pp. WA27–WA39, 2020.
- [26] Y. Yoon, H. G. Jeon, D. Yoo, J. Y. Lee, and I. S. Kweon, "Learning a deep convolutional network for light-field image super-resolution," in *2015 IEEE international conference on computer vision workshop (ICCVW)*, 2015.
- [27] Y.-H. Wang, F.-F. Zhang, J.-J. Liu, C.-J. Xue, and Z.-C. Zhang, "Genesis of the Wurinitu W-Mo deposit, Inner Mongolia, northeast China: constraints from geology, fluid inclusions and isotope systematics," *Ore Geology Reviews*, vol. 94, pp. 367–382, 2018.
- [28] M. D. Zeiler and R. Fergus, "Visualizing and Understanding Convolutional Networks," Cham: Springer International Publishing, 2014.
- [29] X. Zou, L. Zhou, K. Li, A. Ouyang, and C. Chen, "Multi-task cascade deep convolutional neural networks for large-scale commodity recognition," *Neural Computing and Applications*, vol. 32, no. 10, pp. 5633–5647, 2020.
- [30] X. Shi, F. Lv, D. Seng, J. Zhang, J. Chen, and B. Xing, "Visualizing and understanding graph convolutional network," *Multimedia Tools and Applications*, vol. 80, no. 6, pp. 8355–8375, 2021.
- [31] S. Ioffe and C. Szegedy, "Batch normalization: accelerating deep network training by reducing internal covariate shift," in *32nd international conference on machine learning*, Lille, FRANCE, 2015.
- [32] M. Biniz and R. El Ayachi, "Recognition of Tifinagh characters using optimized convolutional neural network," *Sensing and Imaging*, vol. 22, no. 1, p. 28, 2021.
- [33] M. Abadi, A. Chu, I. Goodfellow, et al., "Deep learning with differential privacy," in *23rd ACM conference on computer and communications security (CCS)*, Vienna, AUSTRIA, 2016.
- [34] B. Alabassy, M. Safar, and M. W. El-Kharashi, "A high-accuracy implementation for softmax layer in deep neural networks," in *2020 15th Design & Technology of Integrated Systems in Nanoscale Era (DTIS)*, 2020.
- [35] S. H. S. Basha, S. R. Dubey, V. Pulabaigari, and S. Mukherjee, "Impact of fully connected layers on performance of convolutional neural networks for image classification," *Neurocomputing*, vol. 378, pp. 112–119, 2020.
- [36] D. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *International Conference on Learning Representations*, 2014.
- [37] M. N. Halgamuge, E. Daminda, and A. Nirmalathas, "Best optimizer selection for predicting Bushfire occurrences using deep learning," *Natural Hazards*, vol. 103, no. 1, pp. 845–860, 2020.
- [38] S. J. Reddi, S. Kale, and S. Kumar, "On the convergence of adam and beyond," in *International Conference on Learning Representations, ICLR*, Vancouver, BC, Canada, 2018.
- [39] R. Solgi, H. A. Loáiciga, and M. Kram, "Long short-term memory neural network (LSTM-NN) for aquifer level time series forecasting using in-situ piezometric observations," *Journal of Hydrology*, vol. 601, p. 126800, 2021.
- [40] A. Rusiecki, "Trimmed categorical cross-entropy for deep learning with label noise," *Electronics Letters*, vol. 55, no. 6, pp. 319–320, 2019.
- [41] A. Ji, A. W. Z. Chew, X. Xue, and L. Zhang, "An encoder-decoder deep learning method for multi-class object segmentation from 3D tunnel point clouds," *Automation in Construction*, vol. 137, p. 104187, 2022.
- [42] Y. Liu, W. Li, Z. Feng, Q. Wen, F. Neubauer, and C. Liang, "A review of the paleozoic tectonics in the eastern part of Central Asian Orogenic Belt," *Gondwana Research*, vol. 43, pp. 123–148, 2017.
- [43] L. Wang, K. Qin, G. Song, et al., "Volcanic-subvolcanic rocks and tectonic setting of the Zhengguang intermediate sulfidation epithermal Au-Zn deposit, eastern Central Asian Orogenic Belt, NE China," *Journal of Asian Earth Sciences*, vol. 165, pp. 328–351, 2018.
- [44] C. Zhang, E. Wang, Z. Bi, et al., "Geochronology and isotope geochemistry studies of an epithermal gold deposit in the northern lesser Khingan Range, NE China: the Gaosongshan example," *Ore Geology Reviews*, vol. 105, pp. 356–374, 2019.
- [45] K. Deng, Q. Li, Y. Chen, C. Zhang, X. Zhu, and Q. Xu, "Geochronology, geochemistry and Sr-Nd-Pb-Hf isotopes of the early Jurassic Granodiorite from the Sankuanggou intrusion, Heilongjiang province, Northeastern China: Petrogenesis and geodynamic implications," *Lithos*, vols. 296–299, pp. 113–128, 2018.
- [46] C. Liu, Z. Zhou, G. Wang, et al., "Geochronology and geochemistry of the Late Jurassic bimodal volcanic rocks from Hailisen area, central-southern great Xing'an Range, Northeast China," *Geological Journal*, vol. 53, no. 5, pp. 2099–2117, 2018.
- [47] Z. Zhang, G. Wang, C. Liu, L. Cheng, and D. Sha, "Bagging-based positive-unlabeled learning algorithm with Bayesian hyperparameter optimization for three-dimensional mineral potential mapping," *Computers & Geosciences*, vol. 154, p. 104817, 2021.
- [48] Y. Wang, S. Christley, E. Mjolsness, and X. Xie, "Parameter inference for discretely observed stochastic kinetic models using stochastic gradient descent," *BMC Systems Biology*, vol. 4, no. 1, 2010.
- [49] M. Langer, A. Hall, Z. He, and W. Rahayu, "MPCA SGD—a method for distributed training of deep learning models on spark," *IEEE Transactions on Parallel and Distributed Systems*, vol. 29, no. 11, pp. 2540–2556, 2018.
- [50] J. Duchi, E. Hazan, and Y. Singer, "Adaptive subgradient methods for online learning and stochastic optimization," *Journal of Machine Learning Research*, vol. 12, pp. 2121–2159, 2011.
- [51] C. Traoré and E. Pauwels, "Sequential convergence of AdaGrad algorithm for smooth convex optimization," *Operations Research Letters*, vol. 49, no. 4, pp. 452–458, 2021.

- [52] F.-L. Fan, J. Xiong, M. Li, and G. Wang, “On interpretability of artificial neural networks: a survey,” *IEEE Transactions on Radiation and Plasma Medical Sciences*, vol. 5, no. 6, pp. 741–760, 2021.
- [53] C. Zhao, K.-Z. Qin, G.-X. Song, and G.-M. Li, “Switch of geodynamic setting from the Paleo-Asian ocean to the Mongol-Okhotsk ocean: Evidence from granitoids in the Duobaoshan ore field, Heilongjiang province, northeast China,” *Lithos*, vols. 336–337, pp. 202–220, 2019.