International Journal of Geo-Information

MDPI

Article

Progressive Geological Modeling and Uncertainty Analysis Using Machine Learning

Hong Li [1], Bo Wan [2,3,*], Deping Chu [1], Run Wang [4], Guoxi Ma [5], Jinming Fu [2] and Zhuocheng Xiao [3]

[1] School of Geography and Information Engineering, China University of Geosciences, Wuhan 430078, China
[2] School of Computer Science, China University of Geosciences, Wuhan 430078, China
[3] National Engineering Research of Geographic Information System, Wuhan 430074, China
[4] Geological Environmental Center of Hubei Province, Wuhan 430034, China
[5] Wuhan Zondy Cyber Science & Technology Co., Ltd., Wuhan 430073, China
* Correspondence: wanbo@cug.edu.cn

Abstract: Three-dimensional geological modeling is a process of interpreting geological features from limited sample data and making predictions, which can be converted into a classification task for grid units in the geological space. In sedimentary settings, it is difficult for a single geological classification process to comprehensively express the complex geological spatio-temporal relationships of underground space. In response to this problem, we proposed a progressive geological modeling strategy to reconstruct the subsurface based on a machine learning approach. The modeling work consisted of two-stage classifications. In the first stage, a stratigraphic classifier was built by mapping spatial coordinates into stratigraphic classes, which reflected the geological time information of the geological unit. Then, the obtained stratigraphic class was used as a new feature for the training of the lithologic classifier in the second stage, which allowed the stratigraphic information to be implicitly converted into a new rule condition and enabled us to output the lithologic class with stratigraphic implications. Finally, the joint Shannon entropy of two classifications was calculated to evaluate the uncertainty of the total steps. The experiment built a fine-grained 3D geological model with integrated expression of stratigraphic and lithologic information and validated the effectiveness of the strategy. Moreover, compared with the conventionally trained classifier, the misclassification of the lithologic class between different strata in the progressive classification results has been reduced, with the improvement of the F1-score from 0.75 to 0.78.

Keywords: geological modeling; machine learning; progressive modeling; geological topology; uncertainty

check for updates

Citation: Li, H.; Wan, B.; Chu, D.; Wang, R.; Ma, G.; Fu, J.; Xiao, Z. Progressive Geological Modeling and Uncertainty Analysis Using Machine Learning. *ISPRS Int. J. Geo-Inf.* **2023**, *12*, 97. https://doi.org/10.3390/ijgi12030097

Academic Editors: Wolfgang Kainz and Maria Antonia Brovelli

Received: 7 January 2023
Revised: 20 February 2023
Accepted: 24 February 2023
Published: 26 February 2023

Copyright: © 2023 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

Three-dimensional geological modeling is a key technology for the integration, management, mining, visualization and sharing of geological survey results [1–4]. The main task of 3D geological modeling is constructing a 3D model based on existing geological observations and measurements to reveal spatial structural features, deposition relationships and rock composition in the subsurface [5–7]. It has significant potential for geological environment evaluation, deep mineral exploration, geological disaster prevention and control, and urban underground space development [8–13].

The essence of geological modeling is a process of interpolating and predicting geological features of the whole area of interest from a limited amount of known data [14]. It generally treats geological units as discrete variables, such as stratigraphic classes or rock classes, and thus formulates the modeling task as a classification problem that corresponds to discrete variables [15]. Machine learning is an artificial intelligence method that learns a model with statistical characteristics from known data and uses the model to make judgments and predictions about new scenarios [16,17]. In this way, the principles of 3D geological modeling fit with machine learning algorithms.

In recent years, there has been a boom in research on 3D geological modeling based on machine learning [18–29]. As listed in Table 1, the possibility of various machine learning algorithms in handling geological reconstruction tasks has been verified, from shallow classification algorithms, such as support vector machines (SVM), decision tree (DT), random forest (RF), and maximum likelihood, to variants of neural networks, such deep feedforward neural networks (DFNN), convolutional neural networks (CNN), recurrent neural networks (RNN), graph neural networks (GNN), and generative adversarial networks (GAN). The above studies have established the foundation for the application of machine learning to 3D subsurface reconstruction, but there are relatively few studies on the uncertainty of the prediction results. In addition, they also reveals the characteristics of different approaches. Traditional shallow machine learning classifiers are suitable for learning tasks with small samples, while deep learning algorithms dominated by neural networks generally require large amounts of training data to obtain better learning results. However, it is always troublesome to acquire and label geological data. The samples are too sparse to represent the feature space effectively, thus limiting the performance of deep learning models in geological reconstruction tasks to some extent. Since the application scenario targeted in this work is a modeling task based on borehole core data with a small sample size, the classical shallow machine learning algorithm is more suitable to handle such tasks.

Table 1. Research on 3D geological modeling using machine learning.

Author (Year)	Data	Algorithm	Input	Target	Accuracy	Uncertainty
Smirnoff [18]	Wells, geology map, sections	SVM	Position	Sedimentary facies	79%	-
Wang [19]	Well- logs	SVM	Log parameters	Shale lithofacies	75%	-
Adeli [20]	Drill holes	Decision tree	Position, mineral components	Rock type of iron ore	82.6%	Only mentioned
Xiang [21]	Boreholes, sections, geological map	RF	Strata, rock, gravity and magnetic	Mineralization type	86.84%	-
Gonçalves [22,23]	Wells	Maximum likelihood	Position, orientation	Strata	52.84%	Only mentioned
	Orientation measures	Gaussian process	Position, orientation	Iso-value of potential field	-	Predictive variance
Jia [24]	Borehole (2766) and 3D inversion model	Stacking method	Position, residual density and magnetic susceptibility	Rock types	86%	-
Bai [25]	Training image of flow	CNN	Pixel value	Lithofacies	87.2%	-
Yao [26]	Well-logging, seismic attribute, lithofacies model	DFNN	Position, facies, seismic attribute	Physical properties	-	-
Hillier [15]	Borehole, orientation	GNN	Graph adjacency matrix, graph node matrix	Scalar field, rock	-	-
Zhou [27]	Boreholes	RNN	Coordinates, elevation	Stratum type	62.98%	-
			Coordinates, elevation	Stratigraphic sequence	72.16%	-
Jiang [28]	Land-surface observations, airborne electromagnetic	GAN	Normalized multiple-resolution valley bottom flatness	Paleovalley aquifer index	-	-
Illarionov [29]	Hydrodynamic models of oil fields	End-to-end neural network	Reservoir static variables, initial state, control parameters	Wells' production rates	-	-

Geological modeling based on machine learning often focuses on the prediction of rock types, i.e., lithologic modeling or lithofacies modeling, as shown in Table 1 [15,18–21,24,25].

This may be due to the fact that lithologic assemblage is a reflection of the depositional environment and can also reveal the properties of rocks and soils in the formation, such as color, composition, and unfavorable geological bodies [30,31]. For example, lithologic modeling is particularly important in areas with karst development to detect the potential for geological hazards to occur [32]. To date, most studies used positions and rock properties to predict lithologic types, which successfully reveal the distribution of rocks in 3D space. However, the single lithologic classification does not take into account the complex correlation between lithologic units in geological space and geological time, making it difficult to accomplish modeling work in some specific settings. Especially in sedimentary environments, detailed knowledge of the depositional time and spatial distribution of the sediments is commonly required simultaneously. Stratigraphy states that each geology unit embodies a definite geological history, which can be reconstructed by some specific methods [33–36]. In this case, the stratigraphic information within a 3D geological model implies the geological age information. At this point, it becomes necessary to express lithologic and stratigraphic information synergistically in the same 3D geological model, which can be endowed with temporal and spatial concepts and helps us to understand lithologic characteristics and sedimentary evolution under geological time constraints. Herein, to reconstruct a lithology model with geological time meaning, we proposed a progressive geological modeling strategy using a machine learning algorithm. First, we utilized a machine learning classifier to classify the strata of every grid cell in the subsurface. From the stratigraphic classification results, it can be learnt that the grid units with the same stratigraphic label belong to the sediments of the same geological period. Second, the vectorized stratigraphic label that indicates geological age was introduced into the lithology classifier for lithological classification, and gradually generated a 3D lithology model with geological time significance. Next, we employed the joint Shannon entropy to analyze the uncertainty of the 3D model based on the classification results of the two stages. At the end, the geological spatio-temporal relationships expressed in the resultant geological model were analyzed by using geological topology.

2. Materials and Methods

2.1. Dataset and Preprocessing

The dataset used in this work included 736 geotechnical engineering boreholes in a newly developed urban area in Chengdu, Sichuan, China, containing X, Y, Z coordinates, stratigraphic stratification information, lithologic class information, and geotechnical engineering parameters (bearing capacity characteristic value f_a and internal friction angle φ_i) obtained from borehole core testing.

Since the original samples contained only the stratified points of each borehole, resulting in the sparsity of feature space and samples, resampling of the original borehole data was required. The boreholes were resampled along the *Z*-axis, and a total of 13,641 sample points were obtained. Then, the original boreholes were converted into a series of sample points with spatial position, stratigraphic class, lithologic class, and geotechnical property information. Figure 1 shows the stratigraphy and lithology statistics of the dataset. The pie chart in the upper right reveals the data proportion of the involved four strata, i.e., artificial accumulation (Qh^{ml}), alluvium (Qh^{apl}), Ziyang Formation (Qp^3-Qhz), and Guankou Formation (K_2g), from new to old. All three younger strata are Quaternary sediments, except for the Gunkou Formation, which is a Cretaceous formation. The number of lithologic samples involved in each stratum is shown in the left bar chart.

To eliminate the effect of magnitude, the data were normalized [24,37]. Coordinate values were rescaled to between 0 and 1 using the following linear algorithm:

$$X^* = \frac{x - min}{max - min} \tag{1}$$

where x and X^* represented the values before and after normalization, respectively, and max and min represented the maximum and minimum values, respectively, of every

coordinate. Meanwhile, the two geotechnical features, f_a and φ_i, were processed by the following logarithm function with base 10:

$$Y^* = log_{10} y \quad (2)$$

where y and Y^* denoted the geotechnical values before and after normalization, respectively, of every geotechnical property. Finally, the pre-processed data were divided into a training set (80%) and testing set (20%).

In addition, the geotechnical properties of the non-sample cells were derived from the corresponding property models when predicting the lithology of the entire study area. The model of bearing capacity characteristic values and the model of internal friction angles were provided by the Chengdu Institute of Survey and Investigation (http://www.cdkc.cn/, accessed on 27 October 2021).

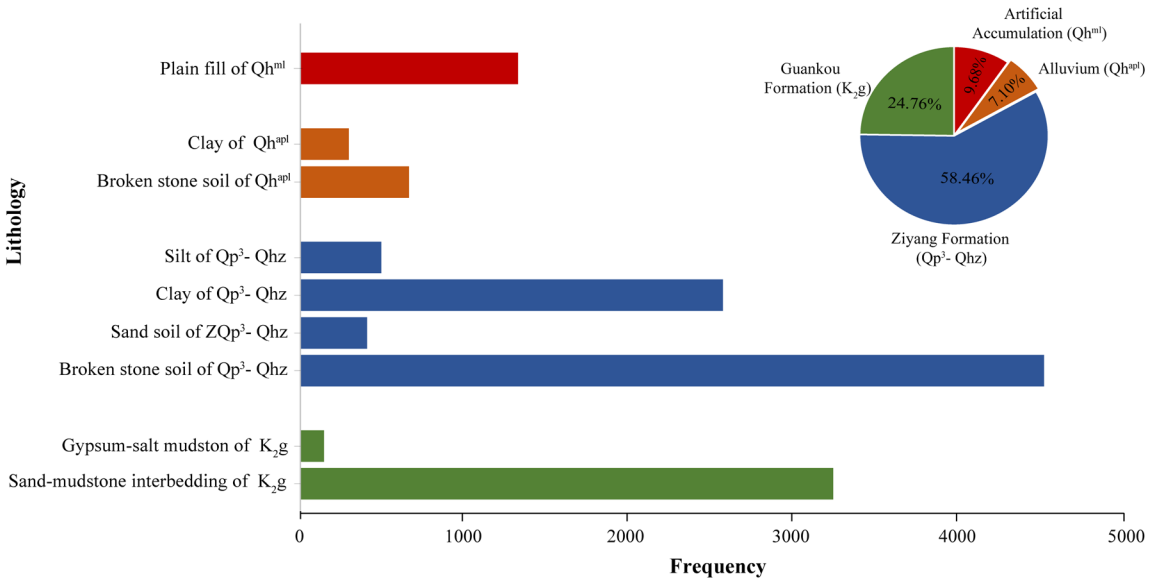


Figure 1. Distribution of stratum and lithology in dataset. The pie chart in the upper right reveals the data proportion of the involved four strata, artificial accumulation, alluvium, Ziyang Formation, and Guankou Formation. The number of lithologic samples involved within each stratum is shown in the left bar chart: (1) artificial accumulation (Qh^{ml}) is in red and dominated by plain fill; (2) alluvium (Qh^{apl}) in orange consists of clay and broken stone soil; (3) the blue Ziyang Formation (Qp^3-Qhz) consists of silt, clay, sand, and broken stone soil; (4) Guankou Formation (K_2g) in green is composed of Cretaceous sediments, including gypsum–salt mudstone and sand–mudstone interbedding.

2.2. Progressive Geological Modeling

The progressive geological modeling framework consisted of two-stage classifications, i.e., the stratigraphic classification and lithologic classification, and classification uncertainty analysis, as shown in Figure 2. The spatial and temporal information contained in the stratigraphic information was employed as a constraint for the latter lithologic prediction. Since random forest (RF) has the advantages of high accuracy, few parameters, and stable performance even with a small amount of data, it fits with the reconstructing task based on small samples [17,38,39]. Therefore, RF was adopted as the classification algorithm. The *Scikit-learn*, a popular opensource Python module for integrating various machine learning algorithms, was used in our experiment [40,41]. In addition, *MapGIS*, a software of geographical information systems, was employed as the visualization tool to display classification results and generate the 3D geological model [42].

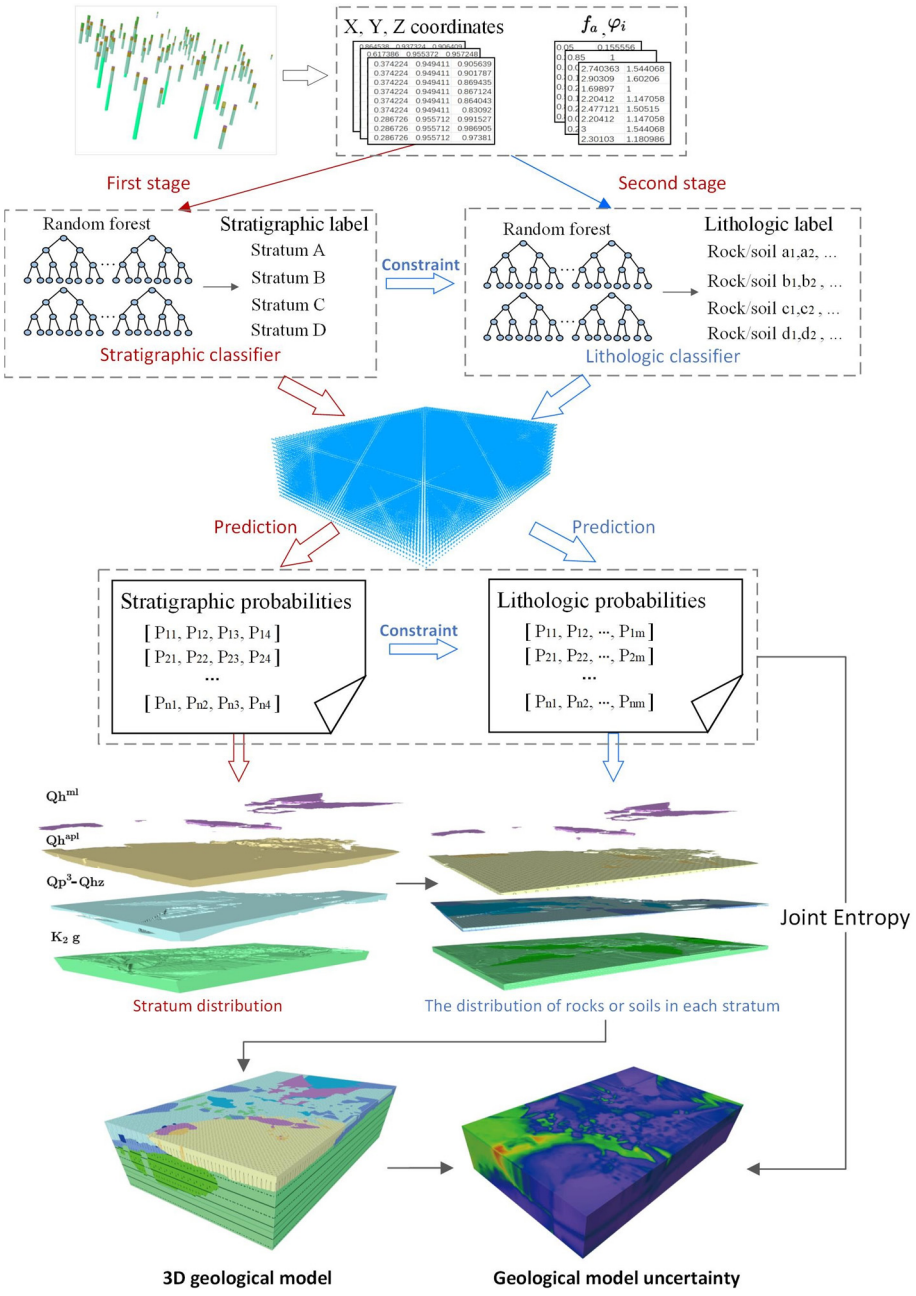


Figure 2. Flowchart of the proposed geological modeling framework in this paper. The modeling work contains stratigraphic classification, lithologic classification, and classification uncertainty evaluation. The path indicated by the red arrows on the left is the stratigraphic classification process by using an RF, and the blue on the right is the lithologic classification process with the constraint of stratigraphic classification results.

(1) Stratigraphic classification

The path indicated by the red arrows on the left in Figure 2 was the classification of the first stage, i.e., the stratigraphic classification process. It took X, Y and Z coordinates as input features and the stratum category as the learning target. The stratum classifier is first trained and tested using training data and test data, and then the trained stratum classifier is used to predict all the 3D grid cells in the reconstruction space. The output of the stratum classifier was an array of probabilities for different stratigraphic classes to which each instance belongs. The stratum type with the highest probability was marked as

the target stratum [43]. Finally, the stratigraphic classification results were imported into the visualization tool to display the stratigraphic distribution.

The size of the 3D grid affects the prediction accuracy, especially in the Z-axis direction. This was mainly because there were thin layers in the study area. If the resolution was too small, i.e., the grid was too large, some thin layers may be missed. Therefore, we chose to increase the resolution, by setting a smaller size in the Z-axis direction, to ensure that each geological category can be predicted. The 3D grid size was set to 5 m × 5 m × 2 m in both classification stages.

(2) Lithologic classification

The path shown by the blue arrows on the right side of Figure 2 was the classification of the second stage, i.e., the lithologic classification process. To further address the subjectivity and complexity in lithologic prediction, geotechnical parameters can be an effective support for lithologic modeling because they provide implicit knowledge of lithologic classes in a high-dimensional vector space. It helps us to discriminate rock units based on their status and physical properties. Thus, the lithologic classification took X, Y, Z coordinates, f_a, φ_i, and the stratigraphic class label from stratigraphic classification as input features. This allowed the stratigraphic prediction results to be used as input for the coupled RF, leveraging the distribution characteristics of the strata to constrain lithologic unit prediction. Unlike stratigraphic classification, the stratigraphic class labels needed to be vectorized before the lithology classifier can be run. This discrete feature was processed here using the one-hot encoding method. Other sub-processes were consistent with stratigraphic classification.

It is important to note that lithologies with the same name but within different strata were identified with different lithologic labels during the lithologic classifying stage. This facilitated the clarification of the lithologic composition and distribution characteristics in each stratum in a 3D space, as this manuscript aimed to learn geological features on a coarse scale to a fine scale through two progressive classifications. The introduction of stratigraphic prediction added a new rule condition to each decision tree in the RF, i.e., the stratigraphic age information was implicitly transformed into a stratigraphic boundary identification vector for lithologic classification. It tried to reduce the possibility that these lithologies with the same name, but different stratigraphic systems, were misclassified. In this case, the results of lithologic classification should be better coupled with the stratigraphic distribution to achieve a finer geological model.

(3) Classification uncertainty evaluation

One of the advantages of using machine learning for geological modeling is the ability to directly quantify the modeling uncertainty using probability array output by the classifier, in contrast to other modeling methods [43]. Information entropy proposed by Shannon is a statistical measure of the dispersion of a random variable [44]. The higher the information entropy, the more discrete the random variable is, and the more uncertain the value of the random variable is. The information entropy reaches its maximum when all uncertainties are present. In the geological classification, the geological unit type of each geological grid cell was considered as a random variable. The information entropy of the geological classification is the expectation of the amount of information from all possible classification events [45]. The stratigraphic classification uncertainty was measured by the Shannon entropy using Equation (3).

$$H[X] = -\sum_{i=1}^{n} P(x_i) log P(x_i) \quad (3)$$

where n corresponded to the number of stratigraphic classes and $P(x_i)$ was the probability of the i-th stratigraphic class.

For the uncertainty of lithologic classification, it not only depended on the probability of the lithologic class, but was also influenced by the probability of the stratigraphic class. That is, the uncertainty of stratigraphic classification results would be transferred to the

subsequent lithologic classification. In this case, we employed joint Shannon entropy to evaluate the uncertainty of the lithologic classification. In fact, it served as a confidence measure of the total classification steps. The joint Shannon entropy was related to entropy and conditional entropy, which both depended on probability distributions [46,47]. It was in the form of Equation (4).

$$H(X,Y) = -\sum_{i=1}^{n}\sum_{j=1}^{m} P(x_i,y_j) log P(x_i,y_j) \tag{4}$$

To further analyze the joint entropy, the mathematical derivation was as follows:

$$\begin{aligned} &H(X,Y)\\ &= -\sum_{i=1}^{n}\sum_{j=1}^{m} P(x_i,y_j) log P(x_i,y_j)\\ &= -\sum_{i=1}^{n}\sum_{j=1}^{m} P(x_i,y_j) log P(x_i) P(y_j|x_i)\\ &= -\sum_{i=1}^{n}\sum_{j=1}^{m} P(x_i,y_j) log P(x_i) - \sum_{i=1}^{n}\sum_{j=1}^{m} P(x_i,y_i) log P(y_j|x_i)\\ &= -\sum_{i=1}^{n} P(x_i) log P(x_i) - \sum_{i=1}^{n}\sum_{j=1}^{m} P(x_i,y_j) log P(y_j|x_i)\\ &= H(X) + H(Y|X) \end{aligned} \tag{5}$$

where X and Y represented the stratigraphic variable and the lithologic variable, respectively, n and m corresponded to the number of stratigraphic classes and lithologic classes, $P(x_i)$ was the probability of the i-th stratigraphic class, $P(x_i,y_j)$ was the joint probability of stratigraphic and lithologic variables, and $P(y_j|x_i)$ represented the conditional probability, which meant the probability that the lithologic class was y_j when the stratigraphic class had been identified as x_i. $H(X)$ represented the single Shannon entropy of the stratigraphic variable and $H(Y|X)$ was the conditional entropy of the lithologic variable when the stratigraphic class was determined.

The joint entropy values were non-negative numbers, and normalized joint entropies were obtained by dividing the entropy value at each position by the maximum of the joint entropies. The normalized joint entropies were then scaled between 0 and 1, with 0 indicating a value of null uncertainty in the classification process and 1 indicating extremely high uncertainty.

3. Results

3.1. Classification Results

Three metrics, including precision, recall, and F1-score, were used to evaluate the performance of the classifiers based on the testing data [48]. As shown in Equations (6)–(8), calculating each metric required the testing results of TP (true positive), TN (true negative), FP (false positive), and FN (false negative) values. All metrics were scaled to be between 0 and 1 and the metrics closer to 1 indicated the better performance of the classifier.

$$precision = \frac{TP}{TP + FP} \tag{6}$$

$$recall = \frac{TP}{TP + FN} \tag{7}$$

$$\text{F1 score} = 2 \cdot \frac{precision \cdot recall}{precision + recall} \tag{8}$$

The testing results of the stratigraphic classification are shown in Table 2, with the overall precision, recall and F1-score values of 0.89. Among the four stratigraphic classes, Qh^{ml} and Qp^3-Qhz obtained higher testing results. Since the data volume of Qp^3-Qhz exceeded half of the dataset, the classifier can fully learn the features of this class, so the best testing results were achieved. However, Qh^{apl} showed contrasting results. Although Qh^{ml} did not dominate the dataset, it achieved better results than expected. By analyzing the spatial distribution of the dataset, it was found that cells of Qh^{ml} were distributed in

the upper part of the reconstruction space; therefore, they were more easily separated from the lower cells.

Table 2. The testing results of the stratigraphic classification.

Stratigraphy Unit	Target Class	Precision	Recall	F1-Score
Qh^{ml}	A	0.93	0.92	0.92
Qh^{apl}	B	0.81	0.83	0.82
Qp^3-Qhz	C	0.96	0.91	0.93
K_2g	D	0.85	0.91	0.87
All units	—	0.89	0.89	0.89

Qh^{ml}: artificial accumulation of Quaternary; Qh^{apl}: alluvium of Quaternary; Qp^3-Qhz: Ziyang Formation of Quaternary; K_2g: Guankou Formation of Cretaceous.

The testing results of the lithologic classification are shown in Table 3, and the overall precision, recall and F1-score values were 0.76, 0.81 and 0.87, respectively. The precision and F1-score values of three lithologies, including class a_1 (plain fill of Qh^{ml}), c_4 (broken stone soil of Qp^3-Qhz), d_2 (sand–mudstone interbedding of K_2g), were 0.9 or above. Similar to the stratigraphic classification, two types of rocks or soils, class a_1 and class d_2, were located at the top and bottom of the reconstructed space, respectively. They were relatively easier to separate from other lithologic units, so better classification results were obtained. Class c_4 had the largest number of samples and achieved good testing results. d_1 (gypsum–salt mudstone of K_2g) had the smallest amount of data among all the lithology categories, whereas the lithologic classifier did not learn enough features, resulting in relatively poor testing performance. This needs further improvement in future research.

Table 3. The testing results of the lithologic classification.

Stratum Unit	Code	Lithologic Unit	Target	Precision	Recall	F1-Score
Qh^{ml}	A	Plain fill	a_1	0.95	0.88	0.91
Qh^{apl}	B	Clay	b_2	0.62	0.65	0.63
		Broken stone soil	b_4	0.78	0.87	0.82
Qp^3-Qhz	C	Silt	c_1	0.60	0.71	0.65
		Clay	c_2	0.87	0.88	0.87
		Sand	c_3	0.88	0.89	0.88
		Broken stone soil	c_4	0.91	0.89	0.90
K_2g	D	Gypsum–salt mudstone	d_1	0.31	0.64	0.42
		Sand–mudstone interbedding	d_2	0.96	0.92	0.94
All units	—	All units	—	0.76	0.81	0.78

Qh^{ml}: artificial accumulation of Quaternary; Qh^{apl}: alluvium of Quaternary; Qp^3-Qhz: Ziyang Formation of Quaternary; K_2g: Guankou Formation of Cretaceous.

To demonstrate the effectiveness of RF in geological classification, we compared the classification results of RF with SVMs, DTs and artificial neural networks (ANN) in the same environment. The parameters of each classification algorithm were set to default values. The target label was the stratigraphic class, and the feature spaces of four classifiers were consistent. As shown in Table 4, the precision, recall and F1-score of RF had certain advantages over the other three algorithms. In the geological classification task, RF outperformed the other three classifiers. This was important since we selected it as the basic classification unit in the proposed modeling framework.

Table 4. Comparison of testing performance of four machine learning algorithms in geological classification.

Classifier	Precision	Recall	F1-Score
SVM	0.8487	0.8537	0.8512
DT	0.8149	0.7967	0.8057
ANN	0.7195	0.7413	0.7302
RF	0.8875	0.8925	0.8850

3.2. Visualization of the Geological Model

The prediction results of the 3D reconstruction space were imported into MapGIS for geological model visualization. The model results and the corresponding cross-sections are shown in Figures 3 and 4, respectively. It should be noted that the values of the latitude and longitude grids have been processed to comply with data privacy rules. We stretched the *Z*-axis 10 times in Figure 4 to more clearly display the geological characteristics. To more intuitively show the distribution of lithology and stratigraphy, colors and textures were combined in the model visualization. Four color families were used to identify four stratigraphic classes, that is, purple for Qh^{ml}, yellows for Qh^{apl}, blues for Qp^3-Qhz, and greens for K_2g. Geological units of the same color family belonged to the same stratum, and geological bodies of the same texture had the same lithology. For instance, the light-yellow broken stone soil of Qh^{apl} and the blue broken stone soil of Qp^3-Qhz were both broken stone soils in lithology terms, so the textures both adopted the symbol of the irregular circle, although they belonged to two different strata. In this way, we obtained a fine-grained geological model. It possessed the ability to obtain the extensional morphology and tip extinction positions of the strata, and provided information on the lithologic composition and distribution within each stratum.

As shown in Figure 5, the lithologic 3D map of each stratum described the lithologic information. We can observe the material components of each stratigraphic layer. For the Qh^{apl}, the geological units of clay were scattered in the formation with poor continuity, while for the Qp^3-Qhz, the units of clay dominated the formation and had good continuity. In addition, in the K_2g, the Gypsum–salt mudstone was interspersed in the sand–mudstone interbedding. In fact, this reflected the geologic genesis. The study area is tectonically part of the West Sichuan Sunken. During the Guankou period of the late Cretaceous, the Chengdu Basin continued to sink due to the Yanshan movement. Under the dry and hot climatic conditions at that time, the saline lake water formed by sea intrusion continuously evaporated and became concentrated. As a result, gypsum-bearing fine-grained sediments were formed. At the beginning of each sedimentary, mudstones and thin carbonates, sandstones and mudstones are rhythmically interbedded, forming the existing lithological distribution characteristics of the Guankou Formation.

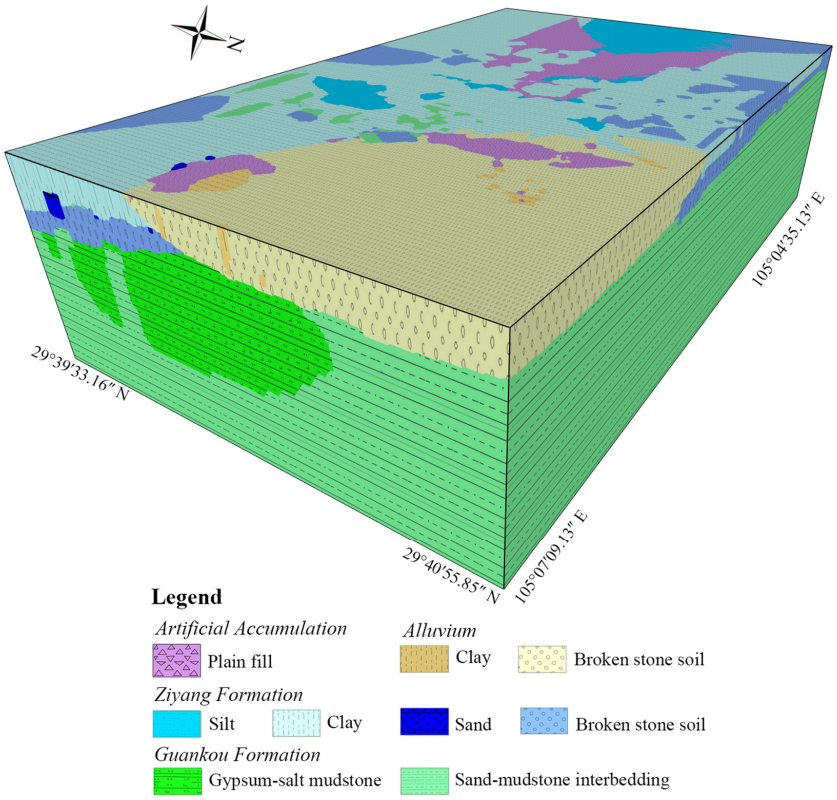


Figure 3. Model result for progressive geological modeling. Color families were used to distinguish strata, and textures were used to mark lithologies. (1) Color family. Purple: artificial accumulation, yellows: alluvium, blues: Ziyang Formation, greens: Guankou Formation. (2) Texture. Triangles: plain fill; vertical dotted lines: clay; circles: broken stone soil; dotted pairs: silt; horizontal lines and single points: sand; horizontal lines and dotted pairs: gypsum–salt mudstone; horizontal dotted lines: sand–mudstone interbedding.

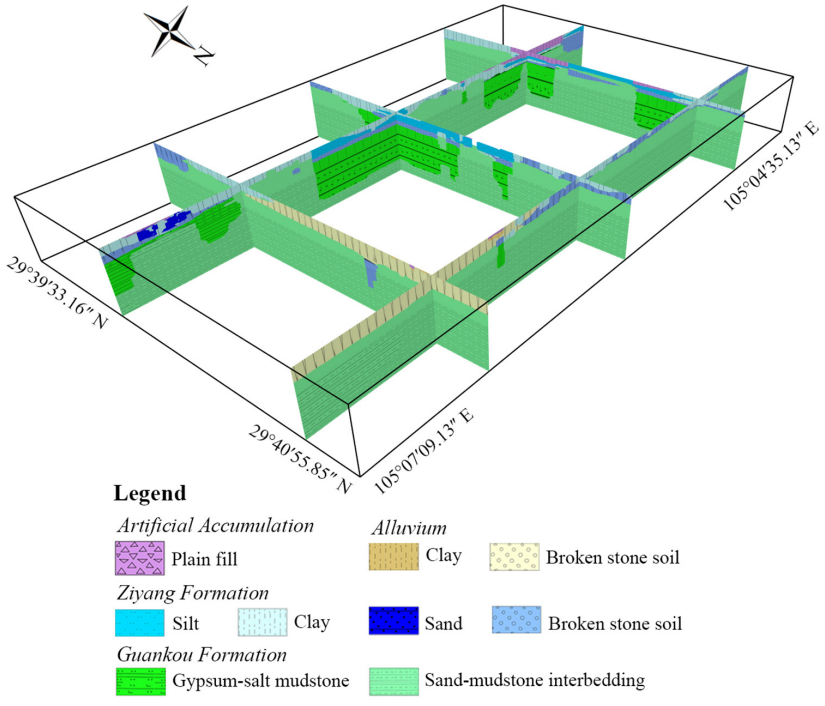


Figure 4. Cross-sections of the 3D geological model.

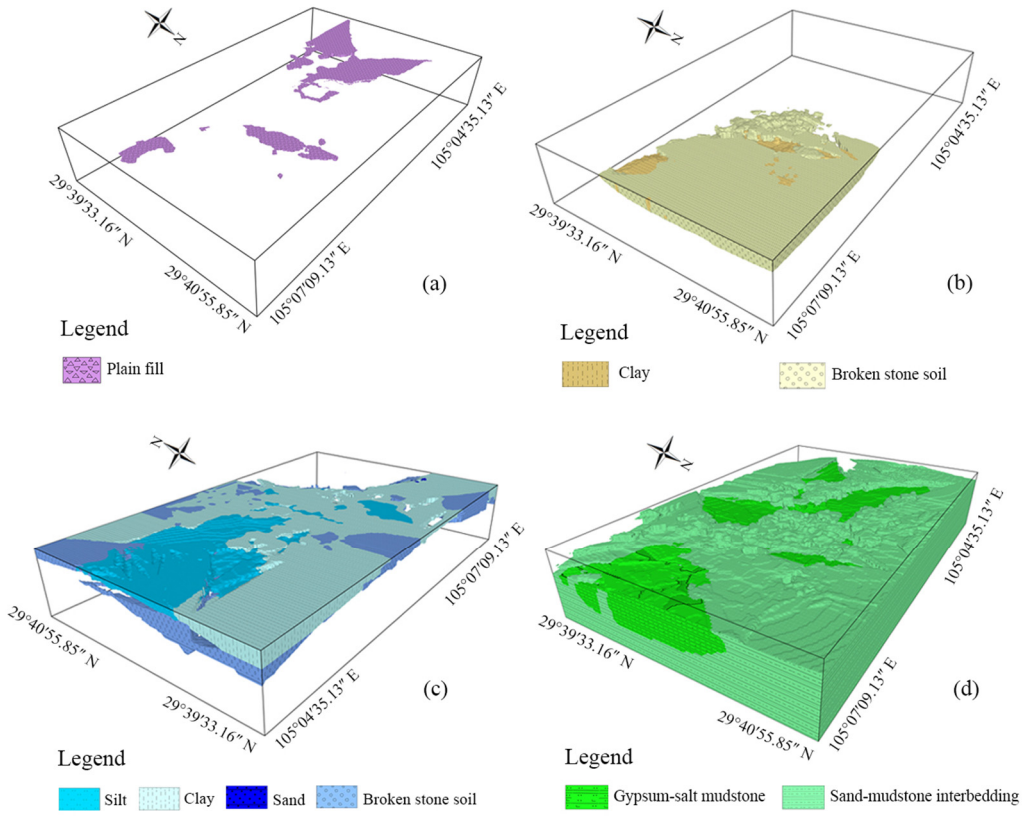


Figure 5. The lithologic 3D map of different stratum in the geological model. (**a**) Qh^{ml}; (**b**) Qh^{apl}; (**c**) Qp^3-Qhz; (**d**) K_2g.

3.3. Modeling Uncertainty

The normalized joint entropy of the results of the two-stage classification represented the uncertainty of the progressive modeling, as shown in Figure 6, with the smallest uncertainty in purple cells and the largest in red. Stratigraphic boundaries tend to demonstrate larger uncertainties, especially interfaces that are adjacent to multiple geological categories. Combined with the geological model in Figure 3, the region with the largest uncertainty was the junction of Qh^{ml}, Qh^{apl} and Qp^3-Qhz. Meanwhile, cells from lithologic boundaries demonstrated higher uncertainty than the lithologically interior cells. This was because cells from geological boundaries had an almost equal likelihood of being classified into all of the contact geological categories, which led to higher information entropy. In progressive modeling, there were multiple classification possibilities for stratigraphy and lithology at the same time because all units possessed dual targets of the stratigraphic label and the lithologic label. Therefore, it was not difficult to explain why the largest uncertainty in the geological model occurred at the stratigraphic boundaries. The uncertainty in the first classification stage was indeed transferred to the lithologic classification stage, resulting in the highest level of confusion at the stratigraphic boundary grid cells.

The average normalized entropy and the cell number statistics for each stratum are displayed in Figure 7, and for lithology, these are displayed in Figure 8. The pie chart on the upper right represents the percentage of cell numbers of each target class in the reconstruction space, and the bar chart displays the average normalized entropy. In Figure 8, it should be noted that the lithologic components of the same stratum were gathered together, using colors to identify the lithologic classes in the bar and pie charts. In contrast to the testing results of stratigraphic classification, the average entropy of Qh^{ml} was the largest, while that of K_2g was the smallest. To explain the unexpected phenomenon, cell counts of each stratigraphic class and lithologic class in the whole reconstructed space were further analyzed. The K_2g dominated the reconstruction space, and most of the K_2g cells were located in the lower area. The cells adjacent to other stratigraphic units accounted for

a very small proportion of the total number of the K_2g, resulting in a low average entropy and low uncertainty. The same was true for the average joint entropy of geological units in progressive modeling (Figure 8); the clay of Qh^{apl} and the sand of Qp^3-Qhz had fewer cells, but higher average joint entropy. This implied that the uncertainty of the 3D geological model was not equivalent to the testing accuracy in the classification process, but rather served as a means to express the likelihood of underground cells being correctly classified through entropy. Moreover, the testing results of the machine learning model were mainly applied to demonstrate the performance of the trained classifier with the testing set and did not fully characterize the prediction results.

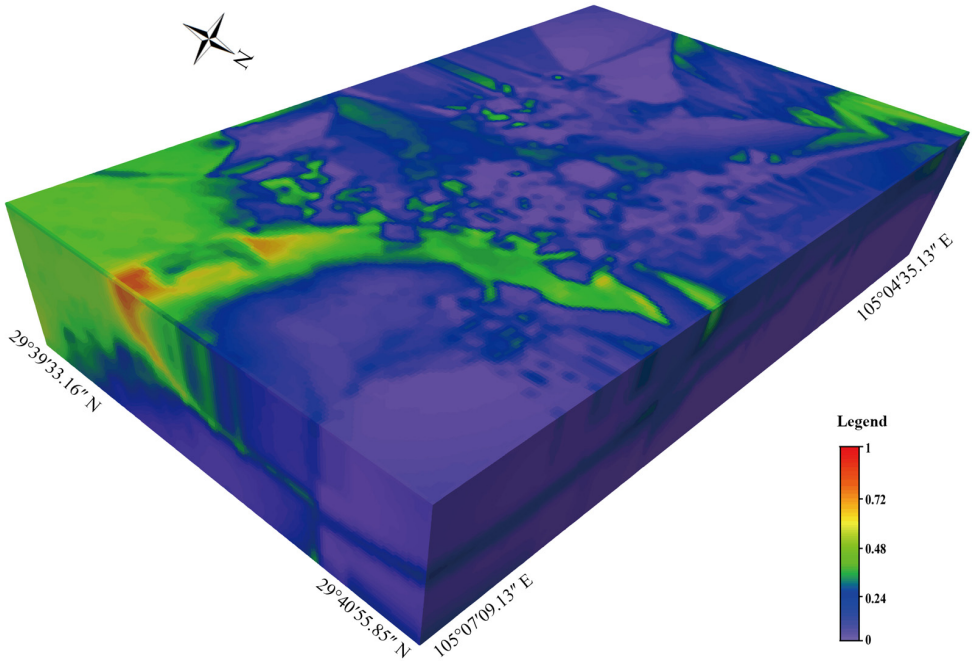


Figure 6. Uncertainty model of progressive modeling.

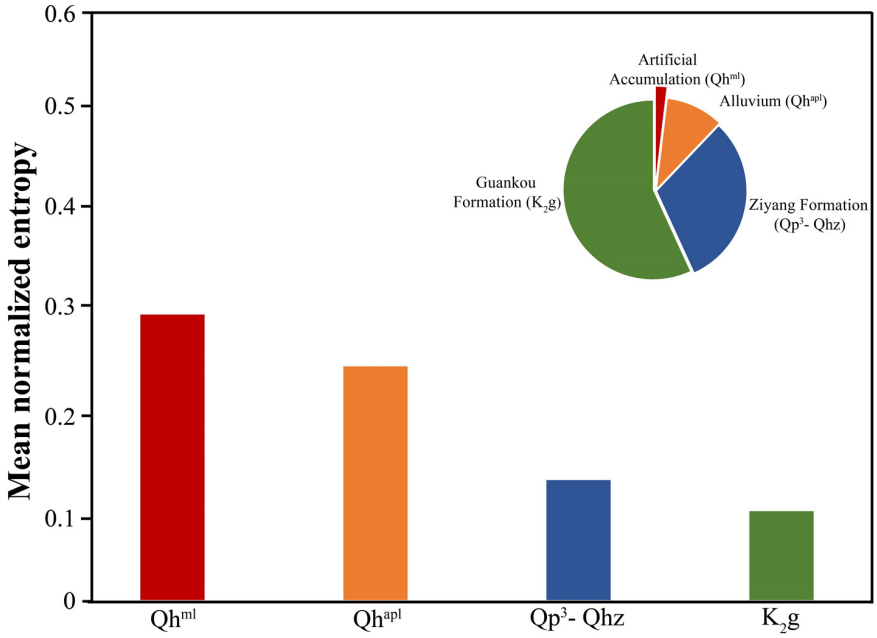


Figure 7. Average normalized entropy and statistics of cell number of each stratigraphic class. The bar chart shows the average normalized entropy of each stratigraphic class. The pie chart on the upper right represents the percentage of cell numbers of each stratigraphic class in the reconstruction space.

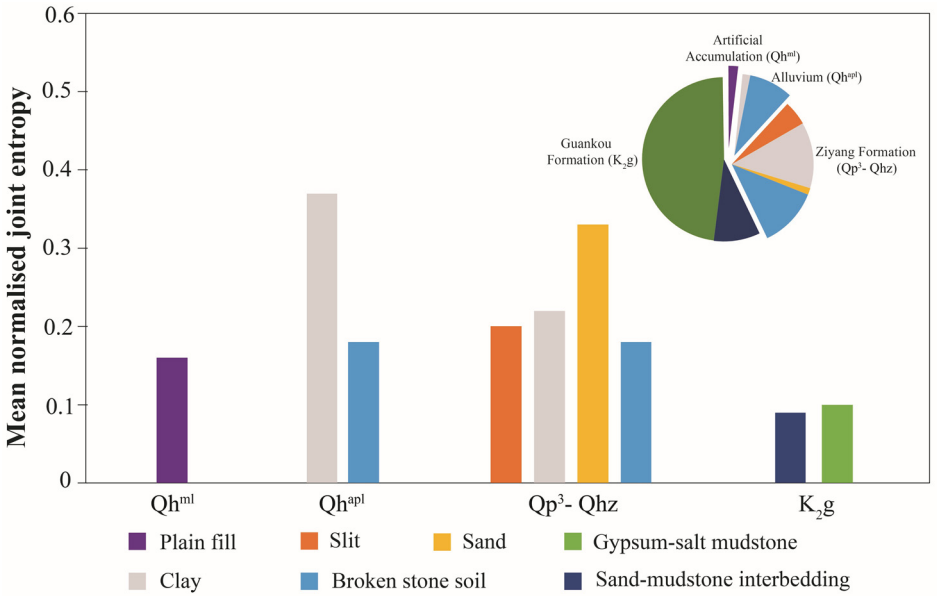


Figure 8. Average value of normalized joint entropy and statistics of cell number of each lithologic class. The bar chart displays the average value of normalized joint entropy. The pie chart on the upper right represents the percentage of cell numbers of each lithologic class in the reconstruction space. Rocks or soils in the same stratum were gathered together, and the different lithologic classes were distinguished by colors.

4. Discussion

4.1. Ablation Study

To investigate the potential of geotechnical properties in geological classification, i.e., f_a and φ_i in this work, several classification results with different input combinations have been analyzed. Input combinations and their corresponding results are shown in Table 5. It was found that group 4, which contained all five features, gained the highest testing success rate. Groups 2 and 3 that only contained one geotechnical property also obtained higher accuracy than group 1, which does not contain geotechnical features. It meant that both geotechnical parameters were effective for geological classification. The f_a represented the pressure value that corresponded to linear deformation in the pressure deformation curve determined by the bearing test and the φ_i was the dip angle of the dislocation plane of rock mass under vertical shear failure, reflecting the internal friction between the particles in the rock. These were significant geotechnical parameters that identified the characteristics of different rocks or soils. It is also worth mentioning that the results of group 1 and group 5, which contained only coordinates and geotechnical properties, respectively, were not particularly poor, indicating that both the location information and geotechnical properties were important for the geoscience research.

Table 5. Classification results of the different input features.

Groups	Input	Target	F1-Score
Group 1	X, Y, Z coordinates	Lithology	0.60
Group 2	X, Y, Z coordinates; f_a	Lithology	0.70
Group 3	X, Y, Z coordinates; φ_i	Lithology	0.62
Group 4	X, Y, Z coordinates; f_a; φ_i	Lithology	0.75
Group 5	f_a; φ_i	Lithology	0.58

4.2. Comparation with the Conventionally Trained Classifier

To verify the effectiveness of the progressive geological modeling strategy, the conventional one-step lithologic classification method was compared. This only included the X, Y, Z coordinates and two geotechnical properties, f_a and φ_i, which were input as classification

features for the prediction of lithologic class. The experimental results are shown in Table 6, and the values of three evaluation metrics in the conventionally trained classifier are 0.73, 0.77, and 0.75. Compared with Table 3, the results of the two-stage progressive lithologic classification with the introduction of stratigraphic information have been improved by about 3 percentage points, and the F1-score was improved from 0.75 to 0.78. The success rates of progressive classification for most of the lithologic classes were increased. Although class d_1 (gypsum–salt mudstone of K_2g) generated poor results in both classifications, the progressive classification results still showed some improvements, especially the recall, which increased from 0.52 to 0.64.

Table 6. Testing results of conventionally lithologic classification.

Stratum Unit	Code	Lithologic Unit	Target	Precision	Recall	F1-Score
Qh^{ml}	A	Plain fill	a_1	0.88	0.88	0.88
Qh^{apl}	B	Clay	b_2	0.48	0.53	0.50
		Broken stone soil	b_4	0.77	0.82	0.79
Qp^3-Qhz	C	Silt	c_1	0.60	0.69	0.64
		Clay	c_2	0.87	0.83	0.85
		Sand	c_3	0.87	0.86	0.86
		Broken stone soil	c_4	0.88	0.87	0.87
K_2g	D	Gypsum–salt mudstone	d_1	0.27	0.52	0.36
		Sand–mudstone interbedding	d_2	0.94	0.91	0.92
All units	—	All units	—	0.73	0.77	0.75

Qh^{ml}: Artificial accumulation of Quaternary; Qh^{apl}: alluvium of Quaternary; Qp^3-Qhz: Ziyang Formation of Quaternary; K_2g: Guankou Formation of Cretaceous.

Considering the imbalance of lithologic classes in the dataset, confusion matrixes were constructed to evaluate the classification results of each lithologic class. Principles about the confusion matrix were obtained from Stehman's work [49]. Normalized confusion matrices of conventional and progressive geological classification are shown in Figure 9a,b. The numbers in red show the misclassification situations in the lithologic classification. They were misclassified as lithologic classes in other strata without exception. However, all of these situations were improved to varying degrees by adding the stratigraphic class label. That is, without the constraint of stratigraphic information, geological units in stratum A may be classified as a rock or soil in stratum B, even if they were two completely different rocks or soils. It reduced the accuracy of lithological modeling. The import of stratigraphic class was capable to increase the classification spacing between lithologic classes of different strata, while the lithologic classes in the same stratum seemed to have no effect. This was understandable because the variability in lithologic classes between different strata increased under the constraint of stratigraphic information. However in the same stratum, it had no effect. Our most interesting finding was the cornerstone of the progressive classification. The constraint information between them was implicitly transformed into a learning feature by using machine learning, which helped us to improve the classification performance of the second stage.

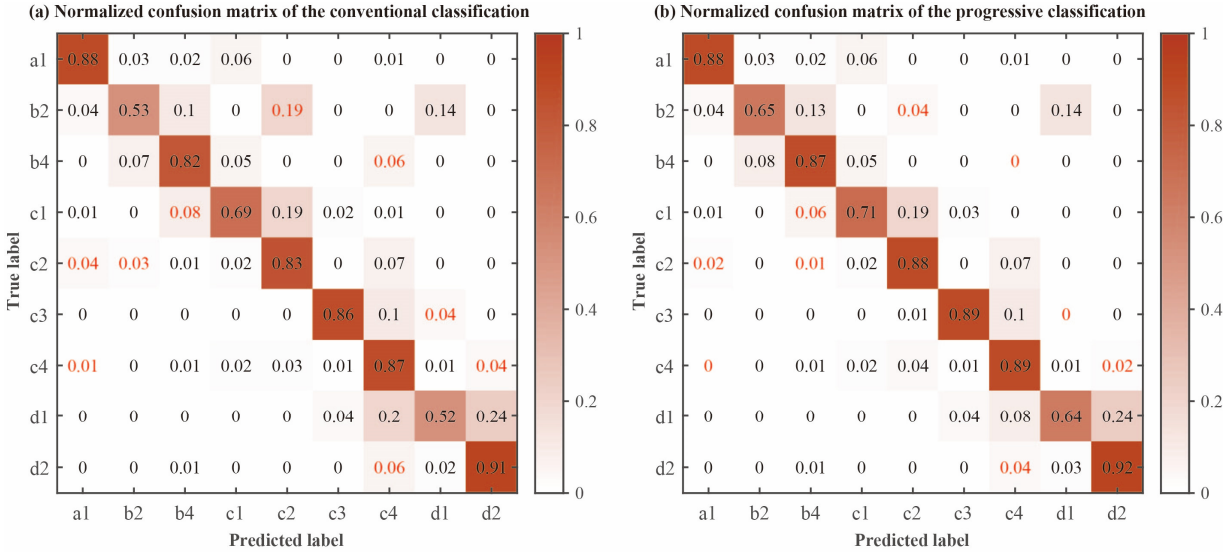


Figure 9. Confusion matrix of the conventional and progressive classification. The words on the axes are interpreted as follows: Qh^{ml}: a1 plain fill; Qh^{apl}: b2 clay; b4 broken stone soil; Qp^3-Qhz: c1 silt; c2 clay; c3 sand; c4 broken stone soil; K_2g: d1 gypsum–salt mudstone; d2 sand–mudstone interbedding.

4.3. Spatio-Temporal Relationships in the Geological Model

The proposed progressive modeling strategy enables the synergistic representation of spatial and temporal relationships among geological elements in the same resultant model. To further elaborate on how the model represented this complicated relationship synchronously, topological analysis of the geology was carried out. In contrast to the conventional definition of topology in geometric graphics, geological topology has its own characteristics. This is because all the topological relationships in the geological field depend on geological processes in the region of interest, such as deposition, intrusion, erosion, and fracture. These historical evolutions transformed the geological topology into a very complex relationship network. Therefore, analyzing the geological topology of the region not only helps to clarify the subsurface structure, but also facilitates the inference of its historical evolutionary processes. We referred to the definition of geological topology in [50], i.e., geological topology was divided into spatial topology and temporal topology, to express the geological topology of the constructed model, as shown in Figure 10.

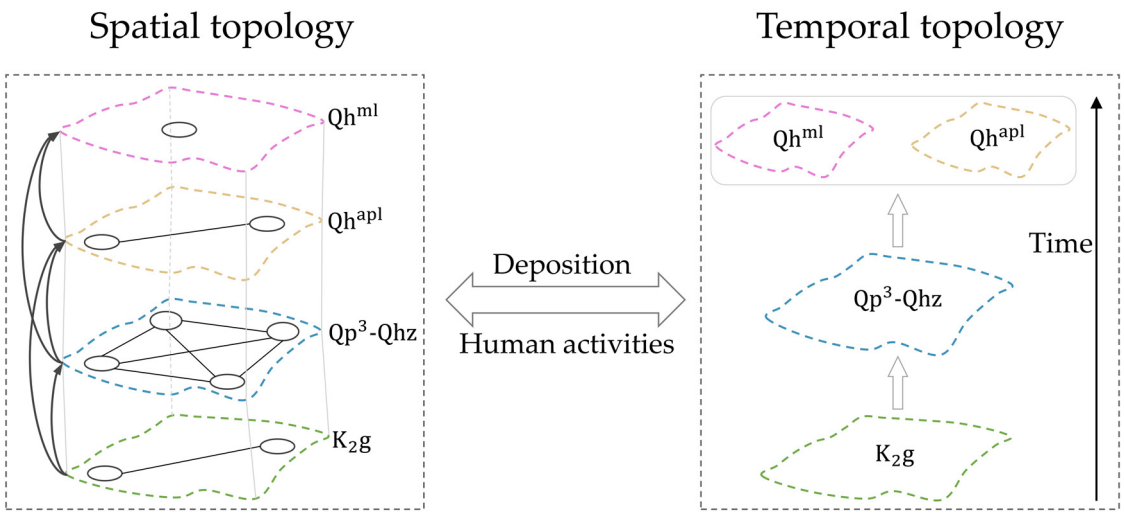


Figure 10. The geological topology of the 3D geological model.

The left part of Figure 10 described the geological spatial topology relationship of the resulting model, which was consistent with the conventional topology defined in geometry. The spatial topology in the geological model is related to the geological geometry and was defined by the Egenhoer relations among the geometric elements [51]. Since there were no obvious fracture structures in the study area of this work, to simplify the spatial topology representation, as suggested by [50], layer-cake stratigraphy was employed. The four irregular polygons with different colors drawn by dotted lines represented the four strata in the model, and the arcs with arrows denoted their contact relations in the geometric space. The circles inside the polygons showed the distribution and contact conditions of lithologic units in each stratum, and the presence of edges between the circles indicated the existence of neighboring relationships between lithologic units. It can be clearly observed that Qh^{ml} was adjacent to two other strata, except for K_2g at the bottom. Qh^{apl} and Qp^3-Qhz were adjacent to three other strata, and K_2g was in contact with Qh^{apl} and Qp^3-Qhz. For the material distribution within each stratum, Qh^{ml} had only one rock type. Qh^{apl} and K_2g were both associated with two lithological categories, while four lithologic materials were spread within the Qp^3-Qhz formation. There were three different structural topologies among multiple stratigraphic layers rock units. It is promising that the generated geological model achieves the fused expression of multi-scale topological information.

In geological semantics, stratigraphic boundaries indicate geological processes that occurred during historical processes, and each geological process implies geological time information. The authors in [52] demonstrated that spatial topology and temporal topology can be inferred from each other by process models, i.e., allowing the transformation of spatial relations into temporal relations. Based on this, the temporal topology of the resulting model was extracted as the right-hand subplot in Figure 10, with arrows pointing to the historical evolutionary direction, i.e., the young geological time direction. It can be found that there were three distinct geological times in the depositional environment of the study area. For simplicity, we noted the geological age of the emergence of K_2g as time 1, Qp^3-Qhz as time 2, and Qh^{ml} and Qh^{apl} as time 3. In the absence of obvious fracture structures and overturned stratigraphy, the old stratum was first formed in the lower part of the space and then overlaid by younger strata. Consequently, it can be clearly deduced from the geological model that the old–young relationships of the three geological times were the oldest at time 1, the second oldest at time 2, and the youngest at time 3. At time 1, because of the presence of gypsum beds in K_2g, the formation of this layer was often accompanied by the end of the sea erosion event in the basin, evaporation and concentration of saline lake water, and deposition of saline material. At time 3, the emerging diluvium and alluvium were transported and accumulated by rivers, and finally deposited into loose sediments to form Qh^{apl}. At the same time, in the process of urban underground space construction, with the transformation, accumulation, filling and other activities of human beings on the subsurface and surface, Qh^{ml} was formed. These geological processes linked the spatial topology with the temporal topology, as shown by the horizontal bidirectional arrow in the center of Figure 10.

From the above analysis, we can conclude that the proposed progressive geological modeling was, to some extent, an integrated expression of stratigraphic and lithologic information in the same model. It allowed the expression of lithologic material distribution, while describing stratigraphic structure information. Meanwhile, the coupling description of spatial information and temporal information was preliminarily realized.

5. Conclusions

A progressive geological modeling framework using a machine learning algorithm has been used to recover the geological structure over the entire volume of interest from discrete and sparse filed data. The framework includes stratigraphic and lithologic classification, supplemented by a joint Shannon entropy calculation of the results of the two stages of classification to quantify the uncertainty of the built model. The research results show that the model allows the expression of lithologic material distribution, while describing the

age of the stratigraphy, which is an integrated representation of stratigraphic and lithologic information in the same model. In addition, by analyzing the geological topology, the model achieves a coupled representation of subsurface spatial and temporal information to a certain extent. In addition, the model's uncertainty shows a trend of low values in the interior of the geological body and high at the edges, following the distribution law of information entropy.

The proposed modeling framework provides a bright future for the construction of multi-grained fine geological models. Moreover, it provides some ideas for model intelligence construction and uncertainty analysis. However, subsurface space is always a field that is not fully understood and needs to be explored continuously to gradually deepen our understanding. With the progression of time, more geological materials and knowledge will be accumulated. Future work should focus on how to incorporate more geological constraints into the geological classification process to improve the classification accuracy and reduce the model uncertainty, so as to generate a more geologically reasonable model.

Author Contributions: Hong Li and Bo Wan contributed the original idea. Guoxi Ma collected the required data. Hong Li, Jinming Fu and Zhuocheng Xiao developed the model and the code used to evaluate the model. Hong Li, Bo Wan, Deping Chu and Run Wang analyzed the results. Hong Li prepared the manuscript with contributions from all co-authors. Bo Wan supervised the entire process. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Chengdu Municipal Bureau of Planning and Natural Resources (Project Number 5101012018002703).

Data Availability Statement: The geological data that supports the experiment of this study cannot be made publicly due to the data use restrictions.

Acknowledgments: The authors thank Ke Liu of China University of Geosciences for their help in model implementation, and Jianglong Xu and Ruirui Yin of Zondy Cyber for their visualization assistance. We also greatly appreciate the data support provided by the Chengdu project team. Lastly, special thanks must be given to the anonymous reviewers for their constructive comments and suggestions, which helped improve our paper.

Conflicts of Interest: The authors declare that they have no known competing financial interest or personal relationships that could have appeared to influence the work reported in this paper.

References

1. Wellmann, F.; Caumon, G. 3-D structural geological models: Concepts, methods, and uncertainties. *Adv. Geophys.* **2018**, *59*, 1–121. [CrossRef]
2. Guo, J.T.; Wu, L.X.; Zhou, W.H.; Jiang, J.Z.; Li, C.L. Towards Automatic and Topologically Consistent 3D Regional Geological Modeling from Boundaries and Attitudes. *ISPRS Int. J. Geo-Inf.* **2016**, *5*, 17. [CrossRef]
3. Calcagno, P.; Chilès, J.P.; Courrioux, G.; Guillen, A. Geological modelling from field data and geological knowledge Part I. Modelling method coupling 3D potential-field interpolation and geological rules. *Phys. Earth Planet. Inter.* **2008**, *171*, 147–157. [CrossRef]
4. Zhang, X.Y.; Zhang, J.Q.; Tian, Y.P.; Li, Z.L.; Zhang, Y.; Xu, L.R.; Wang, S. Urban Geological 3D Modeling Based on Papery Borehole Log. *ISPRS Int. J. Geo-Inf.* **2020**, *9*, 389. [CrossRef]
5. Kemp, E. Spatial Agents for Geological Surface Modelling. *Geosci. Model Dev.* **2021**, *14*, 6661–6680. [CrossRef]
6. Grose, L.; Ailleres, L.; Laurent, G.; Jessell, M. LoopStructural 1.0: Time-aware geological modelling. *Geosci. Model Dev.* **2021**, *14*, 3915–3937. [CrossRef]
7. Linsel, A.; Wiesler, S.; Haas, J.; Baer, K.; Hinderer, M. Accounting for local geological variability in sequential simulations-concept and application. *ISPRS Int. J. Geo-Inf.* **2020**, *9*, 409. [CrossRef]
8. Rienzo, F.; Oreste, P.; Pelizza, S. Subsurface geological-geotechnical modelling to sustain underground civil planning. *Eng. Geol.* **2008**, *96*, 187–204. [CrossRef]
9. He, H.H.; He, J.; Xiao, J.Z.; Zhou, Y.X.; Liu, Y.; Li, C. 3D geological modeling and engineering properties of shallow superficial deposits: A case study in Beijing, China. *Tunn. Undergr. Space Technol.* **2020**, *100*, 103595. [CrossRef]
10. Zhang, Z.; Zhang, J.; Wang, G.; Carranza, E.J.M.; Wang, H. From 2D to 3D modeling of mineral prospectivity using multi-source geoscience datasets, wulong gold district, China. *Nat. Resour. Res.* **2020**, *29*, 345–364. [CrossRef]
11. Leng, X.; Liu, D.; Luo, J.; Mei, Z. Research on a 3d geological disaster monitoring platform based on rest service. *ISPRS Int. J. Geo-Inf.* **2018**, *7*, 226. [CrossRef]

12. Dou, F.F.; Li, X.H.; Xing, H.X.; Yuan, F.; Ge, W.Y. 3D geological suitability evaluation for urban underground space development–A case study of Qianjiang Newtown in Hangzhou, Eastern China. *Tunn. Undergr. Space Technol.* **2021**, *115*, 104052. [CrossRef]
13. Dou, F.F.; Xing, H.X.; Li, X.H.; Yuan, F.; Lu, Z.T.; Li, X.L.; Ge, W.Y. 3D geological suitability evaluation for urban underground space development based on combined weighting and improved TOPSIS. *Nat. Resour. Res.* **2022**, *31*, 693–711. [CrossRef]
14. Frank, T.; Tertois, A.L.; Mallet, J.L. 3D-reconstruction of complex geological interfaces from irregularly distributed and noisy point data. *Comput. Geosci.* **2007**, *33*, 932–943. [CrossRef]
15. Hillier, M.; Wellmann, F.; Brodaric, B.; De Kemp, E.; Schetselaar, E. Three-dimensional structural geological modeling using graph neural networks. *Math Geosci.* **2021**, *53*, 1725–1749. [CrossRef]
16. Karpatne, A.; Ebert-Uphoff, I.; Ravela, S.; Babaie, H.A.; Kumar, V. Machine learning for the geosciences: Challenges and opportunities. *IEEE Trans. Knowl. Data Eng.* **2019**, *31*, 1544–1554. [CrossRef]
17. Rodriguez-Galiano, V.; Sanchez-Castillo, M.; Chica-Olmo, M.; Chica-Rivas, M. Machine learning predictive models for mineral prospectivity: An evaluation of neural networks, random forest, regression trees and support vector machines. *Ore Geol. Rev.* **2015**, *71*, 804–818. [CrossRef]
18. Smirnoff, A.; Boisvert, E.; Paradis, S.J. Support vector machine for 3D modelling from sparse geological information of various origins. *Comput. Geosci.* **2008**, *34*, 127–143. [CrossRef]
19. Wang, G.; Carr, T.; Ju, Y.; Li, C.F. Identifying organic-rich Marcellus Shale lithofacies by support vector machine classifier in the Appalachian basin. *Comput. Geosci.* **2014**, *64*, 52–60. [CrossRef]
20. Adeli, A.; Emry, X.; Dowd, P. Geological modelling and validation of geological interpretations via simulation and classification of quantitative covariates. *Minerals* **2018**, *8*, 7. [CrossRef]
21. Xiang, J.; Xiao, K.; Carranza, E.; Chen, J.P.; Li, S. 3D Mineral Prospectivity Mapping with Random Forests: A Case Study of Tongling, Anhui, China. *Nat. Resour. Res.* **2019**, *29*, 395–414. [CrossRef]
22. Gonçalves, Í.G.; Kumaira, S.; Guadagnin, F. A machine learning approach to the potential-field method for implicit modeling of geological structures. *Comput. Geosci.* **2017**, *103*, 173–182. [CrossRef]
23. Gonalves, T.G.; Guadagnin, F.; Kumaira, S.; Silva, S. A machine learning model for structural trend fields. *Comput. Geosci.* **2021**, *149*, 104715. [CrossRef]
24. Jia, R.; Lv, Y.; Wang, G.; Carranza, E.; Chen, Y.Q.; Wei, C.; Zhang, Z.Q. A stacking methodology of machine learning for 3D geological modeling with geological-geophysical datasets, Laochang Sn camp, Gejiu (China). *Comput. Geosci.* **2021**, *151*, 104754. [CrossRef]
25. Bai, T.; Tahmasebi, P. Hybrid geological modeling: Combining machine learning and multiple-point statistics. *Comput. Geosci.* **2020**, *142*, 104519. [CrossRef]
26. Yao, J.; Liu, Q.; Liu, W.; Liu, Y.Y.; Chen, X.D.; Pan, M. 3D Reservoir Geological Modeling Algorithm Based on a Deep Feedforward Neural Network: A Case Study of the Delta Reservoir of Upper Urho Formation in the X Area of Karamay, Xinjiang, China. *Energies* **2020**, *13*, 6699. [CrossRef]
27. Zhou, C.Y.; Ouyang, J.W.; Ming, W.H.; Zhang, G.H.; Du, Z.C.; Liu, Z. A Stratigraphic Prediction Method Based on Machine Learning. *Appl. Sci.* **2019**, *9*, 3553. [CrossRef]
28. Jiang, Z.J.; Mallants, D.; Gao, L.; Munday, T.; Mariethoz, G.; Peeters, L. Sub3DNet1.0: A deep-learning model for regional-scale 3D subsurface structure mapping. *Geosci. Model Dev.* **2021**, *14*, 3421–3435. [CrossRef]
29. Illarionov, E.; Temirchev, P.; Voloskov, D.; Kostoev, R.; Simonov, M.; Pissarenko, D.; Orlov, D.; Koroteev, D. End-to-end neural network approach to 3d reservoir simulation and adaptation. *J. Pet. Sci. Eng.* **2022**, *208*, 109332. [CrossRef]
30. Dev, V.A.; Eden, M.R. Formation lithology classification using scalable gradient boosted decision trees. *Comput. Chem. Eng.* **2019**, *128*, 392–404. [CrossRef]
31. Sun, J.; Zhang, R.J.; Chen, M.Q.; Li, Q.; Sun, Y.W.; Ren, L.; Zhang, W.G. Real-time updating method of local geological model based on logging while drilling process. *Arab. J. Geosci.* **2021**, *14*, 746. [CrossRef]
32. Wang, Y.; Jing, H.; Yu, L.; Sy, H.J.; Luo, N. Set pair analysis for risk assessment of water inrush in karst tunnels. *Bull. Eng. Geol. Environ.* **2017**, *76*, 1199–1207. [CrossRef]
33. Arab, M.; Belhai, D.; Granjeon, D.; Roure, F.; Arbeaumont, A.; Rabineau, M.; Bracene, R.; Lassal, A.; Sulzer, C.; Deverchere, J. Coupling stratigraphic and petroleum system modeling tools in complex tectonic domains: Case study in the North Algerian Offshore. *Arab. J. Geosci.* **2016**, *9*, 289. [CrossRef]
34. Catuneanu, O. Model-independent sequence stratigraphy. *Earth-Sci. Rev.* **2019**, *188*, 312–388. [CrossRef]
35. Yu, Y.X.; Xia, Z.M. Study on the application of seismic sedimentology in a stratigraphic-lithologic reservoir in central Junggar Basin. In Proceedings of the 3rd International Conference on Advances in Energy, Environment and Chemical Engineering, Chengdu, China, 26–28 May 2017; Volume 69, pp. 1–7. [CrossRef]
36. Milad, B.; Slatt, R.; Fuge, Z. Lithology, stratigraphy, chemostratigraphy, and depositional environment of the Mississippian Sycamore rock in the SCOOP and STACK area, Oklahoma, USA: Field, lab, and machine learning studies on outcrops and subsurface wells. *Mar. Pet. Geol.* **2020**, *115*, 104278. [CrossRef]
37. Zavadskas, E.; Turskis, Z. A New Logarithmic Normalization Method in Games Theory. *Informatica* **2008**, *19*, 303–314. [CrossRef]
38. Cracknell, M.; Reading, A. 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]

39. Merembayev, T.; Kurmangaliyev, D.; Bekbauov, B.; Amanbek, Y. A Comparison of Machine Learning Algorithms in Predicting Lithofacies: Case Studies from Norway and Kazakhstan. *Energies* **2021**, *14*, 1896. [CrossRef]
40. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
41. Raschka, S.; Liu, Y.; Mirjalili, V. *Machine Learning with PyTorch and Scikit-Learn*; Packt Publishing Ltd.: Birmingham, UK, 2022; p. 101.
42. Yuan, Y.; Shao, C.F.; Ji, X.; Xiang, H.K.; Zhang, W.J. True 3D Surface Feature Visualization Design and Realization with MapGIS K9. In Proceedings of the 7th International Conference on Green Intelligent Transportation System and Safety, Nanjing, China, 1–4 July 2016; Volume-419, pp. 13–27. [CrossRef]
43. Fuentes, I.; Padarian, J.; Iwanaga, T.; Vervoort, R.W. 3D lithological mapping of borehole descriptions using word embeddings. *Comput. Geosci.* **2020**, *141*, 104516. [CrossRef]
44. Shannon, C.E. A Mathematical Theory of Communication. *Bell Systems Tech. J.* **1948**, *27*, 623–656. [CrossRef]
45. Wellmann, J.F.; Regenauer-Lieb, K. Uncertainties have a meaning: Information entropy as a quality measure for 3-D geological models. *Tectonophysics* **2012**, *526*, 207–216. [CrossRef]
46. Ji, W.P.; Wu, J.J.; Zhang, M.; Liu, Z.Z.; Shi, G.M.; Xie, X.M. Blind Image Quality Assessment with Joint Entropy Degradation. *IEEE Access.* **2019**, *7*, 30925–30936. [CrossRef]
47. Liu, Y.; Zheng, Z.; Zhao, L.; Wang, Z. Quality assessment of post-consumer plastic bottles with joint entropy method: A case study in Beijing, China. *Resour. Conserv. Recycl.* **2021**, *175*, 105839. [CrossRef]
48. Powers, D. Evaluation: From precision, recall and f-measure to roc, informedness, markedness & correlation. *J. Mach. Learn. Technol.* **2011**, *2*, 37–63.
49. Stehman, S. Selecting and interpreting measures of thematic classification accuracy. *Remote Sens. Environ.* **1997**, *62*, 77–89. [CrossRef]
50. Thiele, S.; Jessell, W.M.; Lindsay, M.; Ogarko, V.; Wellmann, J.F.; Pakyuz-Charrier, E. The topology of geology 1: Topological analysis. *J. Struct. Geol.* **2016**, *91*, 27–38. [CrossRef]
51. Egenhofer, M.J.; Herring, J.R. Categorizing binary topological relations between regions, lines and points in geographic databases. In *The 9-Intersection: Formalism and Its Use for Natural-Language Spatial Predicates*; National Center for Geographic Information and Analysis: Buffalo, NY, USA, 1994.
52. Burns, K.L. Retrieval of tectonic process models from geologic maps and diagrams. In Proceedings of the Meeting of Geoscience Information Society, Cincinnati, OH, USA, 2–5 November 1981; pp. 105–111.

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.