

Article

A Generalized Borehole Autoregressive Neural Network for 3D Geological Modeling

Hao Li, Chi Zhang and Zhenwen He * 

School of Computer Science, China University of Geosciences (Wuhan), 388 Lumo Road, Wuhan 430074, China; 1202321557@cug.edu.cn (H.L.); 1202211228@cug.edu.cn (C.Z.)

* Correspondence: zwhe@cug.edu.cn

Abstract

Three-dimensional geological modeling is a fundamental technology for reconstructing subsurface geological structures and plays an important role in resource exploration, disaster prediction, and engineering construction. With increasing energy demand and growing environmental safety challenges, accurate characterization of the morphology and physical properties of subsurface strata has become essential for the efficient development of underground space. Machine learning-based three-dimensional geological modeling methods using borehole data reformulate the modeling process as a stratum classification task, thereby reducing manual intervention and improving the level of automation in geological modeling. In this process, the classification of stratigraphic spatial points is a key step, as its accuracy directly influences the quality of the resulting geological body model. However, traditional algorithms typically rely solely on spatial coordinate features to determine stratum affiliation. Such a single-feature-driven approach has limited capability in representing the true morphology of subsurface strata. To address this limitation, this paper proposes a stratum classification method based on Vertical Alignment–Horizontal Distance Weighting (VA-HDW), which is designed to capture spatial correlations between strata and boreholes. On this basis, a specialized neural network model, termed the Generalized Borehole Autoregressive Neural Network (GBARNN), is designed and trained to improve the classification performance of stratigraphic spatial points, thereby contributing to improved three-dimensional geological body modeling quality.

Keywords: machine learning; 3D geological modeling; stratum classification; neural network

1. Introduction

3D geological modeling is a process that utilizes computer technology to represent geological data in a 3D spatial context, simulating the morphology, structure, and attribute distribution of geological bodies in 3D space. As a core technical method in fields such as mineral resource exploration [1], groundwater resource management [2], and geological hazard prediction [3], it provides intuitive references for geological decision-making. Geological drilling is associated with high costs. How to efficiently utilize limited borehole data to improve exploration efficiency and enhance the quality of exploration results has become a major and urgent problem to be solved in the current field of geological engineering. Relying on scientific stratigraphic division technology can not only conduct in-depth analysis of the characteristics of underground geological structures but also provide an accurate and reliable original basis for 3D geological modeling.



Academic Editor: Igor Aizenberg

Received: 27 November 2025

Revised: 7 January 2026

Accepted: 3 February 2026

Published: 5 February 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

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

conditions of the [Creative Commons](https://creativecommons.org/licenses/by/4.0/)

[Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

Early three-dimensional (3D) geological modeling primarily relied on the digitization of geological cross-sections or geometric interpolation of borehole data. Researchers reconstructed fault surfaces and stratigraphic interfaces by standardizing cross-section data and extracting key feature points [4], or sequentially constructed boundaries, faults, and rock bodies by following the stratigraphic depositional order [5]. In addition, for large-scale mining areas, complex geological sequences and geotechnical properties can be jointly modeled by using multiple intersecting cross-sections combined with a “layer-to-solid” transformation approach [6]. Although cross-section-based methods can reasonably and realistically represent geological bodies with faults, they typically involve extensive manual intervention and suffer from low updating efficiency. In contrast, borehole-based methods achieve a higher degree of automation by employing techniques such as Delaunay triangulation [7], surface-based modeling theories [8], or geostatistical simulations (e.g., PACSIM [9]). Models constructed from borehole data can achieve relatively high accuracy and smooth, geologically reasonable structures; however, their performance is strongly constrained by the spatial sparsity of borehole data, geological complexity, and the limited number and high acquisition cost of boreholes. Consequently, there is an urgent need for more innovative and effective modeling approaches.

To overcome the computational limitations of traditional methods, researchers have introduced machine learning algorithms for stratigraphic attribute prediction. Methods such as support vector machines (SVMs) [10,11], improved k-nearest neighbor (KNN) algorithms [12], and relevance vector machines (RVMs) [13] have demonstrated strong robustness in small-sample classification tasks and are capable of effectively capturing the nonlinear characteristics of stratigraphic interfaces. Although these approaches improve modeling efficiency, their input features are often limited to simple spatial coordinates, failing to fully exploit the implicit stratigraphic ordering constraints embedded in borehole data. As a result, their generalization capability in data-sparse regions remains limited.

In recent years, deep learning has become a major research focus in complex geological feature recognition due to its strong feature extraction capability. Convolutional neural networks (CNNs) have been employed to extract structural information from seismic data [14], while generative adversarial networks (GANs) have been used to generate spatial distributions of sedimentary bodies consistent with geological rules, effectively alleviating the limitation of single-pattern representations in traditional sequential simulation methods [15]. By introducing multi-scale generative frameworks [16] or integrating multi-source geophysical data [17], and by exploiting Transformer models to capture long-range dependencies in borehole sequences [18], these approaches have further improved the reliability of 3D geological modeling under sparse borehole constraints. Despite the progress achieved under various conditions, significant challenges remain when dealing with sparse data, complex stratigraphy, or multi-source data fusion. To address the problem of geological data sparsity, the semi-supervised SDLP algorithm proposed in [19] constructs pseudo-labels using a triangulated irregular network (TIN), thereby alleviating the shortage of training samples. In [20], a maximum-likelihood implicit modeling framework was introduced to achieve global joint modeling without manually specifying variogram functions. In addition, ref. [21] employed data self-organization, autoencoders, and ResCapsNet models to significantly enhance the representation accuracy of geological bodies under sparse borehole data conditions.

Based on current research, three-dimensional geological modeling faces accuracy challenges in stratigraphic spatial point classification. Given the complex distribution of subsurface spaces, relying solely on three-dimensional spatial coordinate features for classification inevitably leads to issues such as weak generalization, low accuracy, and poor reliability. As a critical step in three-dimensional geological body modeling, the

effectiveness of stratigraphic spatial point classification methods directly impacts the final modeling quality.

To address these bottlenecks, this paper proposes a stratigraphic classification method based on Vertical Alignment—Horizontal Distance Weighting (VA-HDW). This approach moves beyond sole reliance on 3D coordinate features by constructing a geological spatial field model that explicitly characterizes the horizontal neighborhood effects and vertical similarity between spatial points and their neighboring drill holes. Based on this feature, this paper further designs a GBA-RNN neural network model. Through experimental validation on real urban geological datasets, this method significantly improves stratum classification accuracy in sparse data environments, providing effective decision support for constructing highly reliable 3D geological models.

2. Materials and Methods

2.1. Feature Construction of Geological Spatial Field

2.1.1. Stratigraphic Spatial Point Upsampling

Borehole data typically includes key information such as borehole coordinates (X, Y), borehole elevation, stratum thickness, stratum base depth, borehole number, and borehole stratum number. Since the original data exists in a continuous one-dimensional form and the number of digits varies significantly among different data items, it cannot be directly used for neural network training. Therefore, it is necessary to perform upsampling on the borehole data before training to ensure the uniformity and applicability of the data.

Due to the significant differences in thickness among various strata in boreholes, the number of samples from thicker strata is usually much higher than that from thinner strata. If equal-interval sampling is adopted, the balance of the data may be affected—this causes sample data from thicker strata to dominate during network training, which in turn easily leads to incorrect prediction of the stratigraphic attributes in unknown areas, similar to those of thicker strata. Moreover, as the difference in stratum thickness increases, the probability of misclassification also rises accordingly. To address this issue, this study adopts a non-equal-interval sampling method (see Figure 1). By adjusting the sampling interval according to the thickness of different strata, the balance of sampled data is ensured to a certain extent, thereby improving the accuracy of prediction.

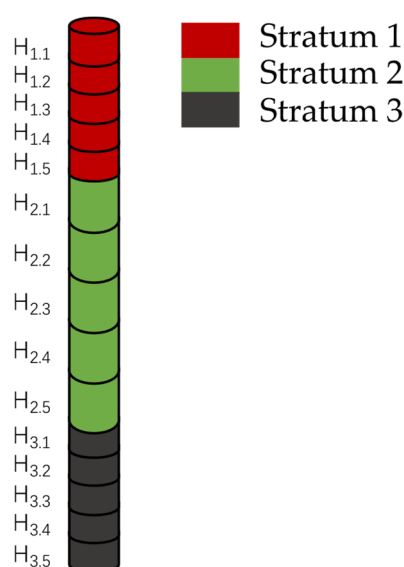


Figure 1. Borehole Sampling.

As shown in Figure 1, the stratigraphic data are displayed in a strip shape and distributed continuously in the vertical direction. Within the depth interval of each stratum, its attributes are continuous and unique, with no data gaps between different strata, H_{ij} representing the j -th part of the stratum i . The calculation formula for the sampling interval D_{ij} of the j -th layer in the i -th borehole is as follows:

$$D_{ij} = \left| \frac{d_{ij} - d_{ij-1}}{n} \right| \quad (1)$$

Among them, d_{ij} represents the base elevation of the j -th layer in the i -th borehole, and n denotes the number of samples per stratum.

2.1.2. Spatial Point Feature Construction

A geological spatial field can be regarded as a complex system composed of various geological spatial points, which takes the geological features of each spatial point as its main characteristics. These features are integrated to form a complete geological information expression system. Taking spatial point a as an example, the determination of its features depends on its association with the borehole set A . The borehole set A contains n borehole samples, and each borehole sample exerts an influence on the geological attributes of the spatial point. Specifically, the features of a spatial point mainly include three aspects: first, the horizontal distance between the spatial point and each borehole, which reflects the spatial proximity between the spatial point and the borehole; second, the corresponding stratum number, which indicates the stratum position where the spatial point is located; third, the influence coefficient, which is used to quantify the contribution degree of each borehole to the geological features of the spatial point a .

The formula for the horizontal distance between the stratigraphic spatial point a and the borehole i is as follows:

$$d_{ai} = \sqrt{(x_a - x_i)^2 + (y_a - y_i)^2} \quad (2)$$

Among them, d_{ai} denotes the horizontal distance between the spatial point a and the i -th borehole, and (x_a, y_a) and (x_i, y_i) represent the planar coordinates of the spatial point a and the i -th borehole, respectively.

The formula for the stratum number corresponding to the spatial point a and the borehole i is as follows:

$$\Phi_{ai} = \{k | top_k \leq z_a \leq bottom_k, k \in U\} \quad (3)$$

Among them, Φ_{ai} denotes the stratum number corresponding to the spatial point a and the borehole i , which starts counting from 1. top_k represents the top elevation of the k -th layer, $bottom_k$ denotes the base elevation of the k -th layer, z_a is the z -coordinate of the spatial point a , and U stands for the set of strata contained in the borehole i .

The calculation formula for the influence coefficient of borehole i on spatial point a is as follows:

$$S_{ai} = F_c \left(\frac{\min\{\Phi_{ai}^s - z_a, z_a - \Phi_{ai}^e\}}{\Phi_{ai}^s - \Phi_{ai}^e} \right) \quad (4)$$

Among them, S_{ai} represents the influence coefficient of the borehole i on the spatial point a , Φ_{ai}^s denotes the top elevation of the stratum Φ_{ai} , Φ_{ai}^e represents the base elevation of the stratum Φ_{ai} , and $F_c(x)$ is a coefficient conversion function. To reduce the impact of complex terms on the experiment, this paper selects the linear function $y = 2 * x$ as $F_c(x)$ to measure the influence coefficient in a simple and intuitive form, and it exactly maps the influence coefficient to between $[0, 1]$.

Considering that the orifice elevation and maximum depth contained in different boreholes may vary, there is an issue where the depth of the spatial point a does not fall within the depth range of a borehole. To ensure the rationality and continuity of the attribute calculation for spatial points outside the borehole range, upper and lower bounds are imposed on Φ_{ai} .

There is a borehole b , whose stratum sequence from top to bottom contains a total of n layers, denoted as $[S_1, S_2, \dots, S_n]$. The orifice elevation is top , and the hole bottom elevation is bot . For the spatial point O with coordinates (x, y, z) , if $z > top$ or $z < bot$, then Φ_{Ob} is calculated by the following formula:

$$\Phi_{Ob} = \begin{cases} S_1, & z > top \\ S_n, & z < bot \end{cases} \quad (5)$$

And at this time, the influence coefficient is assigned according to the minimum value, i.e., $S_{Ob} = 0$.

As shown in Figure 2, it reflects the schematic diagram of obtaining the horizontal distance d , the corresponding stratum number Φ , and the influence coefficient S for the spatial point O from three boreholes.

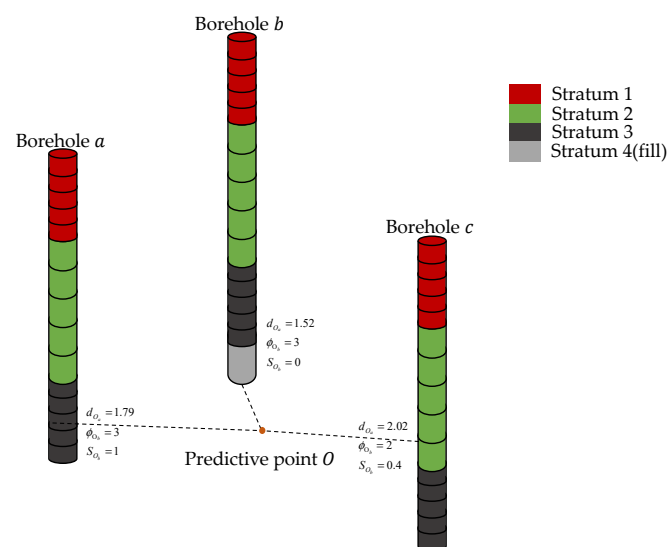


Figure 2. Acquisition of Corresponding Stratum Numbers and Influence Coefficients.

Through the above three feature construction methods, the comprehensive feature of Borehole i for Spatial Point a can be expressed as:

$$Fea_{ai} = (d_{ai}, \Phi_{ai}, S_{ai}) \quad (6)$$

Assuming there are a total of n boreholes within the study area, n features can be constructed for Spatial Point a , and their feature matrix form is as follows:

$$\begin{bmatrix} d_{a1} & \Phi_{a1} & S_{a1} \\ d_{a2} & \Phi_{a2} & S_{a2} \\ \vdots & \vdots & \vdots \\ d_{an} & \Phi_{an} & S_{an} \end{bmatrix} \quad (7)$$

The construction process of this geological spatial field is shown in Figure 3.

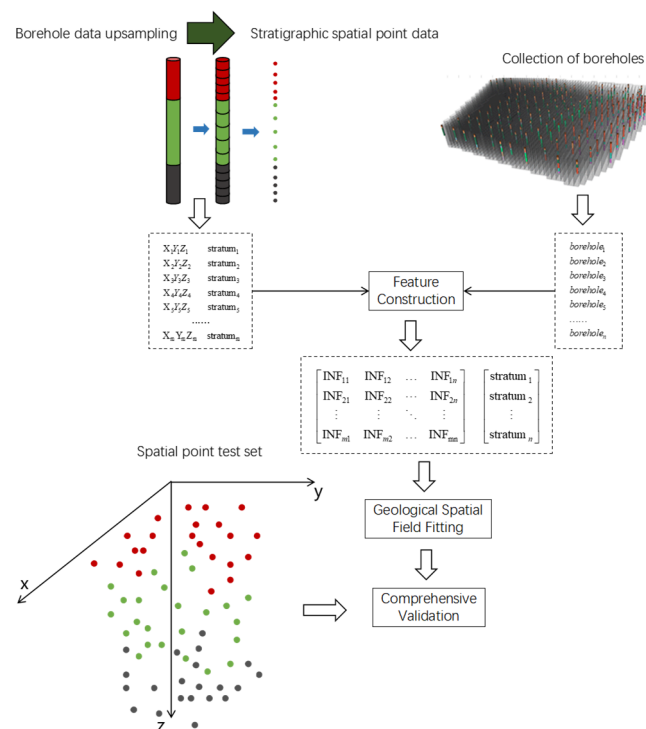


Figure 3. Geological Spatial Field Construction Diagram.

2.2. Neural Network Architecture GBARNN Supporting VA-HDW Stratigraphic Classification Method

2.2.1. Definition of Local Model Unit GBDNN

In the previous section, this paper constructs three key features of stratigraphic spatial points through directional construction. These features are used to build the geological spatial field, thereby implementing the VA-HDW stratigraphic classification method. The essence of this method is to comprehensively measure the influence degree of each borehole on the specified stratigraphic spatial point, and finally determine the stratigraphic attribute of this spatial point. For ease of understanding, this paper proposes the generalized spatial point-borehole distance to characterize the comprehensive influence degree of these three features.

Given a spatial point a and a borehole b , the generalized spatial point-borehole distance D between a and b is expressed by the following function:

$$D_{ab} = F(d_{ab}, \Phi_{ab}, S_{ab}) \quad (8)$$

To accurately depict the nonlinear distribution of elements in 3D stratigraphic space, this study designs a Generalized Borehole Distance Neural Network (GBDNN) unit model. This model integrates feature information of spatial points and boreholes across multiple dimensions, such as horizontal distance, vertical coordinates, and burial depth, to dynamically generate generalized borehole distance parameters. The GBDNN can not only efficiently decompose complex nonlinear relationships but also achieve adaptive representation of the multidimensional coupling correlation between spatial points and boreholes, providing more accurate technical support for geological spatial analysis and modeling. The generalized spatial point-borehole distance can be simply expressed through this unit as:

$$D_{ab} = F(d_{ab}, \Phi_{ab}, S_{ab}) = \text{GBDNN}(d_{ab}, \Phi_{ab}, S_{ab}) \quad (9)$$

Through a networked training mechanism, the GBDNN unit can automatically extract and generate generalized spatial point-borehole distance data, accurately representing the

complex attribute characteristics of specific spatial elements. This module dynamically captures the connection characteristics between strata and boreholes in multi-dimensional space, integrating key parameters such as horizontal distance, corresponding stratigraphic horizon, and influence coefficient, thus achieving efficient modeling and expression of the nonlinear evolution process of strata in 3D space. The structure of GBDNN is shown in Figure 4.

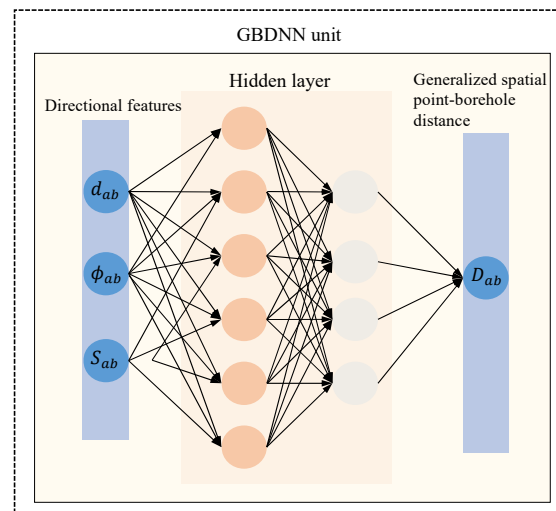


Figure 4. Structural Diagram of the Generalized Borehole Distance Neural Network (GBDNN) Mode.

2.2.2. Definition of the GBARNN Model

In spatial regression modeling, the spatial weight value is used to quantify the relative influence of different boreholes on the stratigraphic attribute prediction of a target spatial point. A larger spatial weight indicates that the corresponding borehole is closer to the prediction point in terms of spatial location and stratigraphic structure, and thus contributes more significantly to the prediction; conversely, a smaller weight implies a weaker influence of that borehole on the target point. Based on the above definition, let the point to be predicted be a , and the borehole set B contain n pieces of borehole data. Based on the idea of spatial regression (This reflects the core interpretation of the term “Autoregressive” in the GBARNN model, and does not refer to the autoregressive mechanism in conventional time-series analysis.), this study constructs a spatial weight matrix by calculating the generalized spatial point-borehole distance between point a and each borehole in the set B . Obviously, the spatial weight value between the point to be predicted and the borehole set is related to its generalized distance to the borehole set. To further characterize the complex characteristics of the geological spatial field, it is also necessary to establish a nonlinear relationship between the spatial weight value λ and the generalized spatial point-borehole distance. By introducing a nonlinear mapping function, the value of λ can be dynamically adjusted to more accurately reflect the correlation strength between the spatial point and the borehole, thereby establishing a spatial correlation model between point a and the borehole set B .

The non-linear function between the defined spatial weight vector λ_i and the generalized spatial point-borehole distance D is as follows:

$$\lambda_a = (w_{a1}, w_{a2}, \dots, w_{an}) = f(D_{a1}, D_{a2}, \dots, D_{an}) \quad (10)$$

Among them, λ_a can be interpreted as a vector composed of the spatial weight values of each borehole in the borehole set B for the prediction point a . w_{ai} is the spatial weight value between the i -th borehole and the prediction point a , and D_{ai} is the generalized

distance between the i -th borehole and the prediction point. Since the data source is borehole data, to prevent the problem of self-fitting, a correction coefficient k is set to update the calculation formula of w_{ai} , so that the spatial weight value between the prediction point to be predicted and its own borehole is set to 0, as follows:

$$w_{ai} = w_{ai} \times k_{ai} \quad (11)$$

Among them:

$$k_{ai} = \begin{cases} 1, & a \text{ is not located on borehole } i \\ 0, & a \text{ is located on borehole } i \end{cases} \quad (12)$$

That is:

$$w_{ai} = \begin{cases} f(D_{ai}), & a \text{ is not located on borehole } i \\ 0, & a \text{ is located on borehole } i \end{cases} \quad (13)$$

Combining the GBDNN unit described in the previous section, the initial spatial weight p_{ai} obtained through the training of the generalized borehole distance neural network is multiplied by the correction coefficient k_{ai} to obtain the final weight w_{ai} between the two points. Given m prediction points and n boreholes, its matrix form is as follows:

$$\begin{aligned} w = p * k &= \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ p_{m1} & p_{m2} & \cdots & p_{mn} \end{bmatrix} \begin{bmatrix} k_{11} & k_{12} & \cdots & k_{1n} \\ k_{21} & k_{22} & \cdots & k_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ k_{m1} & k_{m2} & \cdots & k_{mn} \end{bmatrix} \\ &= \begin{bmatrix} p_{11}k_{11} & p_{12}k_{12} & \cdots & p_{1n}k_{1n} \\ p_{21}k_{21} & p_{22}k_{22} & \cdots & p_{2n}k_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ p_{m1}k_{m1} & p_{m2}k_{m2} & \cdots & p_{mn}k_{mn} \end{bmatrix} \end{aligned} \quad (14)$$

It can be seen from the form of w that to obtain the final result of the prediction point, it is necessary to multiply the weight matrix by a matrix with exactly n rows to obtain the data information related to the predicted stratum. From the perspective of results, to obtain the stratum attribute of the prediction point, it is actually required to calculate the probability distribution of the prediction point in each stratum, and the one with the largest probability value is taken as the final prediction result. To better reflect the stratum probability distribution characteristics of spatial points, in this calculation process, this paper incorporates the stratum thickness feature, namely, the stratum thickness distribution matrix of boreholes. This matrix not only provides the thickness information of boreholes in different strata but also transmits the spatial influence of boreholes to the prediction point through the combination with spatial weights, realizing accurate prediction of stratum attributes.

Let the stratum thickness distribution matrix of boreholes be b , with a dimension of $n \times z$, where n represents the number of boreholes and z represents the total number of strata in the study area. Each row of matrix b corresponds to a borehole, and each column corresponds to the thickness value of a certain stratum, thereby comprehensively describing the thickness distribution of each borehole in different strata. The form of the stratum thickness distribution matrix is as follows:

$$b = \begin{bmatrix} stratum_{11} & stratum_{12} & \cdots & stratum_{1z} \\ stratum_{21} & stratum_{22} & \cdots & stratum_{2z} \\ \vdots & \vdots & \ddots & \vdots \\ stratum_{n1} & stratum_{n2} & \cdots & stratum_{nz} \end{bmatrix} \quad (15)$$

where $stratum_{ij}$ represents the thickness ratio of the j -th stratum in the i -th borehole, and there is:

$$\sum_{j=1}^z stratum_{ij} = 1, i = 1, 2 \dots n \quad (16)$$

By multiplying the weight matrix w (with a shape of $m \times n$, representing the spatial influence weight values between m prediction points and n boreholes) by matrix b , a result matrix of $m \times z$ can be obtained. Each row in this matrix represents a prediction point, and each column represents the weighted thickness value of a certain stratum. These weighted thickness values can be further converted into a probability distribution through normalization processing, thereby reflecting the attribute probability of the prediction point in different strata.

To summarize, the model structure of GBARNN is shown in Figure 5.

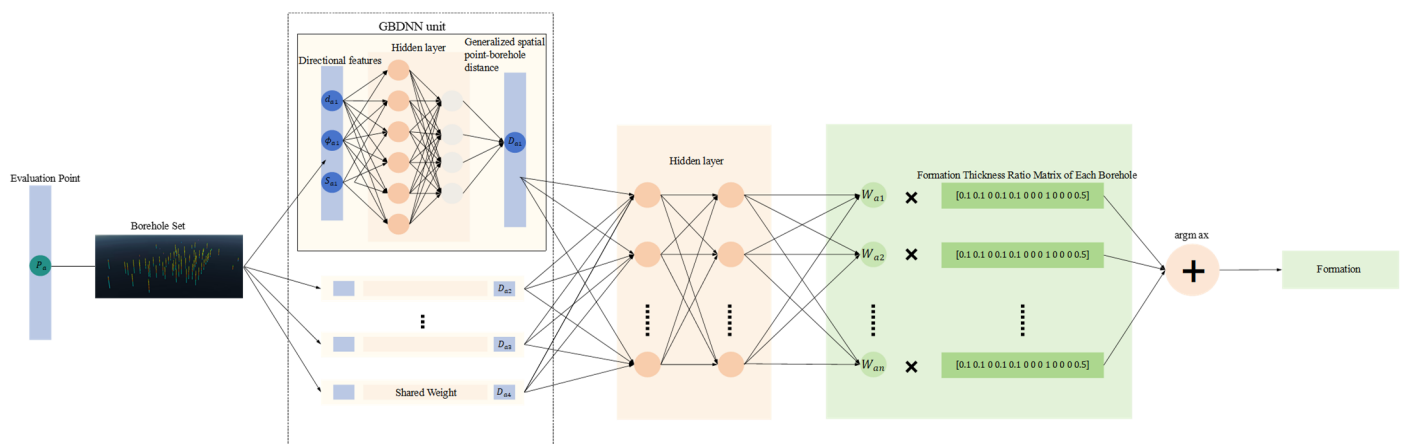


Figure 5. Generalized Borehole Autoregressive Neural Network (GBARNN) Model Structure Diagram.

During the modeling process, the generalized distance components from the prediction points to the boreholes in 3D space are input into the GBDNN units. All GBDNN units share network weights and bias parameters, and through the training process, the network can output the generalized spatial distance between spatial points and boreholes under the geological spatial context. This distance serves not only as the output of GBDNN but also as the input value of GBARNN. In GBARNN, after nonlinear calculations in the hidden layer, the output layer generates a spatial weight matrix. To further optimize the weight distribution, this paper multiplies each element of this matrix by the corresponding element of the correction coefficient matrix to obtain the final spatial weight matrix. This matrix is then multiplied by the stratum thickness ratio matrix of boreholes to generate a stratum attribute probability distribution vector for the prediction points. By finding the maximum value in the probability distribution vector, the stratum attribute value of the prediction points can be determined.

Due to the lack of recognized true values of generalized spatial point-borehole distance as supervision signals during the training process, the GBDNN unit cannot be trained independently. It can only be embedded into the neural network-based overall framework and participate in its training and calculation processes. This design enables the collaborative optimization of GBDNN with GBARNN, allowing it to gradually learn reasonable representations of generalized distance and provide more accurate spatial weight information for stratigraphic attribute prediction.

2.2.3. Adaptability Analysis of VA-HDW to Stratigraphic Pinch-Outs

Stratigraphic pinch-out refers to a geological phenomenon in which a stratum gradually thins and eventually disappears under the influence of depositional or tectonic

processes. It commonly occurs in sedimentary strata and represents an important form of stratigraphic termination. Stratigraphic pinch-outs typically develop in response to changes in the depositional environment, a reduction in sediment supply, or tectonic deformation. Figure 6 illustrates a typical example of a stratigraphic pinch-out phenomenon.

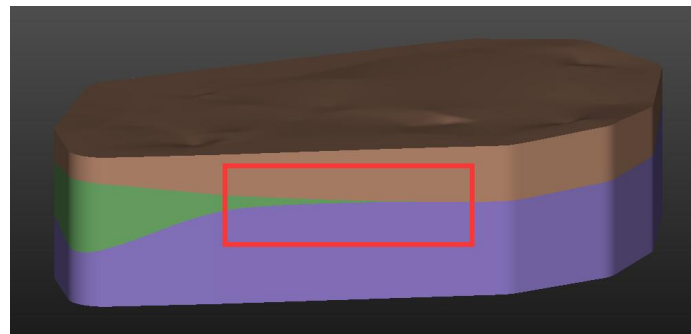


Figure 6. Stratigraphic pinch-out phenomenon.

The VA-HDW method integrates vertical alignment features and horizontally distance-weighted information, enabling stratigraphic pinch-out modeling to a certain extent. Specifically, vertical alignment (VA) is first applied to align the predicted spatial point with boreholes in the vertical direction, allowing the relative burial depth of the spatial point within each borehole to be calculated. This process is crucial for accurately capturing pinch-out characteristics and helps avoid misjudgments that may arise when relying solely on absolute depth. After vertical alignment, VA-HDW further introduces a horizontal distance weighting (HDW) strategy by computing the horizontal distances between the target prediction point and surrounding boreholes. Different weighting factors are assigned to boreholes such that those closer to the prediction point exert a greater influence, while more distant boreholes contribute less. On this basis, VA-HDW employs a deep learning framework to fuse stratigraphic thickness information and outputs a stratigraphic probability distribution matrix with high reliability. When the thickness of a stratum gradually decreases with increasing distance, the trends computed by VA-HDW can be used to determine whether the stratum pinches out at a given location.

As illustrated in Figure 7, during the prediction of stratigraphic classes at spatial points, the influence of different boreholes on the prediction point gradually changes as the point moves from left to right. Initially, the prediction point is mainly influenced by Boreholes 1, 2, and 3, but this influence gradually weakens with increasing distance. Meanwhile, as the prediction point approaches Boreholes 4 and 5, their influence correspondingly increases. This transition process facilitates a gradual shift of the stratigraphic assignment of the spatial point from Stratum B to Stratum A, thereby improving the rationality of the prediction results.

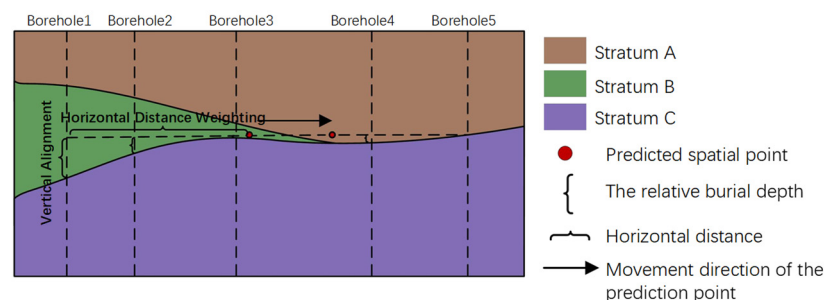


Figure 7. Adaptability Analysis of VA-HDW to Stratigraphic Pinch-Outs.

2.2.4. Workflow of the VA-HDW Algorithm

The core idea of the VA-HDW algorithm is to construct a generalized distance between stratum spatial points and boreholes by jointly considering horizontal distance and vertical anisotropy, and to optimize the spatial weight distribution using a neural network, thereby achieving high-accuracy prediction of the stratigraphic attributes at target points. The workflow of the algorithm is illustrated in Figure 8.

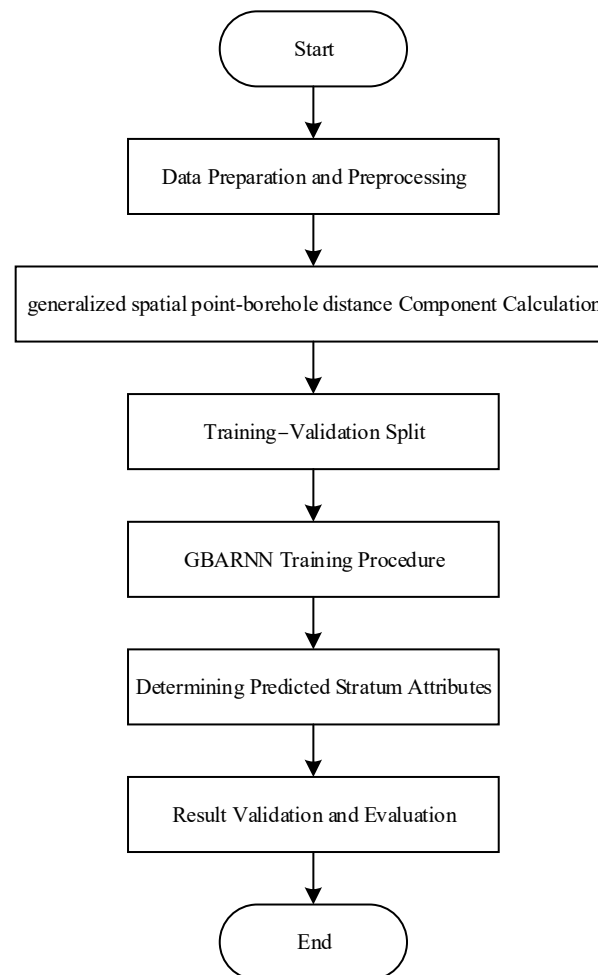


Figure 8. Flowchart of the VA-HDW Algorithm.

(1) Data Preparation and Preprocessing

Borehole data are collected, including borehole coordinates (X , Y), collar elevation, stratum thickness, and stratum attributes. These data are normalized to ensure that different dimensions share a consistent scale. Based on the stratum thickness information of each borehole, a borehole–stratum thickness matrix b ($n \times z$) is constructed, where n denotes the number of boreholes and z denotes the number of strata.

(2) Training and Test Dataset Partitioning

The set of borehole spatial points is divided into training and test datasets according to a specified ratio. Random sampling is adopted to ensure that both datasets are representative in terms of stratum attribute distribution.

(3) Computation of Generalized Point–Borehole Distance Components

For all borehole spatial points, the generalized distance components between each point and all boreholes are computed. These components include horizontal distance

(Euclidean distance) and vertical alignment information (relative burial depth and the corresponding stratum). Through this step, a sample dataset for GDARNN is generated with a shape of $m \times n \times 3$, where m is the number of prediction points and n is the number of boreholes.

(4) GDARNN Model Training

The generalized point–borehole distance components of the training set are fed into the GDDNN network unit. The GDDNN unit is jointly trained with GBARNN in an end-to-end manner to optimize model parameters. Cross-entropy loss is used to measure the discrepancy between predicted and true stratum attributes, and network weights are updated via backpropagation.

(5) Generation of the Spatial Weight Matrix

The generalized spatial distances produced by the GDDNN network unit are input into GBARNN, where nonlinear transformations in the hidden layers generate an initial spatial weight matrix. This initial matrix is then adjusted by incorporating a correction coefficient matrix to obtain the final spatial weight matrix a ($m \times n$), where m is the number of prediction points and n is the number of boreholes.

(6) Computation of Stratum Attribute Probability Distribution

The spatial weight matrix a is multiplied by the borehole–stratum thickness matrix b to obtain the stratum attribute probability distribution matrix ($m \times z$) for the prediction points. The probability distribution for each point is then normalized to yield the probability values across different strata.

(7) Determination of Predicted Stratum Attributes

For each prediction point, the stratum with the maximum probability in its attribute distribution is selected as the final prediction. The spatial coordinates of the prediction point and its corresponding predicted stratum attribute are then output.

(8) Result Validation and Evaluation

The model performance is comprehensively evaluated using the validation dataset of stratum spatial points. Key metrics such as prediction accuracy, recall, and F1-score are computed to quantitatively assess the model's performance in stratum attribute prediction.

2.3. Intelligent 3D Geological Modeling Method

To achieve 3D geological body modeling, the target area is first subjected to spatial discretization based on the preset grid resolution, dividing it into regular 3D grid cells that serve as the basic framework for stratum attribute prediction. The boundaries of the modeling area are determined in a borehole data-driven manner: the spatial coordinates of all stratum spatial points are extracted, and their minimum bounding box is calculated to ensure the modeling range fully covers the borehole distribution area. On this basis, stratum classification prediction is performed based on the spatial coordinates of the grid cells, constructing a 3D geological body model with spatial consistency and data integrity. The minimum bounding box of stratum spatial points is shown in Figure 9.

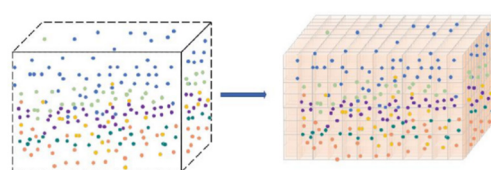


Figure 9. Minimum Bounding Box of Stratum Spatial Points.

After establishing the spatial bounding box, to enhance the constraint of stratum surfaces, several “detection boreholes” need to be set. The trained meta-classifier is then used to predict the stratum horizon for each depth segment of these boreholes. By analyzing the variation in stratum horizons in each “detection borehole”, stratum boundary points can be extracted (the relevant effect is shown in Figure 10), and each stratum surface is constructed by fitting based on the Kriging interpolation method.

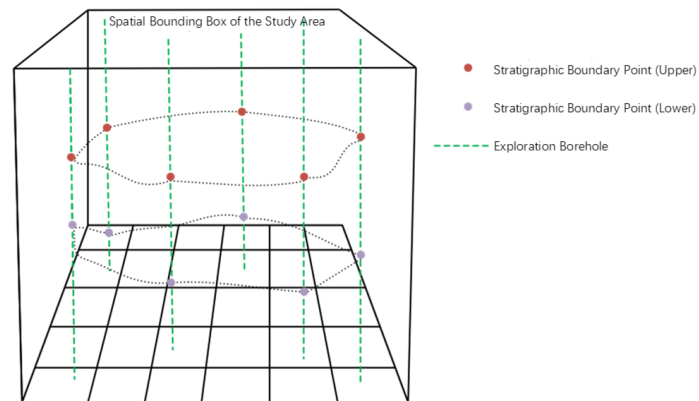


Figure 10. Constraint of Upper and Lower Stratum Boundary Points.

On this basis, the region between two adjacent stratum surfaces is regarded as a complete stratum unit. Under the constraint of the spatial bounding box, the stratum horizon attribute of each grid cell is calculated by combining the Marching Cubes algorithm, thereby constructing a complete 3D geological body model.

3. Results

3.1. Experimental Data

The experimental data used in this paper include the geological exploration borehole dataset. The borehole dataset contains key information such as borehole ID, borehole coordinates, hole mouth elevation, stratum bottom burial depth, and stratum number, providing basic support for geological body modeling. After removing outliers and invalid data, the borehole dataset includes a total of 64 borehole records, covering 7 strata, which are sequentially numbered 1 to 7 from top to bottom. The spatial distribution of the boreholes is shown in Figure 11.

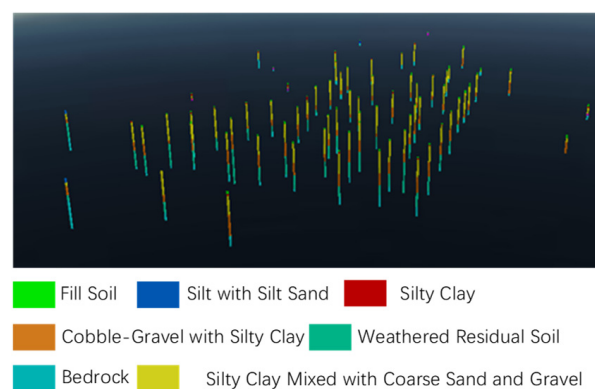


Figure 11. Spatial distribution of boreholes.

Partial borehole data information is shown in Table 1.

Table 1. Partial Borehole Data Information.

X-Coordinate	Y-Coordinate	Borehole ID	Hole Mouth Elevation	Borehole Depth
123,700.0	135,243.4	1	7.43	49
123,538.0	135,249.2	2	7.47	49.8
123,933.3	135,134.7	3	8.2	52.2
123,716.8	135,076.8	4	8.9	65.6

Based on the stratum bottom burial depth parameter in the borehole data, a stratum thickness matrix can be formed. This matrix participates in the training of the GBARNN network, providing a direction for convergence. Taking the sample with Borehole ID 1 as an example, its corresponding stratum thickness data are shown in Table 2. These data accurately reflect the thickness differences between various rock strata, providing important reference information for subsequent stratum division. Through systematic analysis and comprehensive evaluation of the stratum thicknesses of numerous boreholes, the spatial distribution laws and evolution characteristics of the underground geological structure can be fully analyzed, thereby offering scientific support for accurate stratum division.

Table 2. Stratum Thickness Information.

Borehole ID	Stratum Number	Stratum Thickness
1	1	3.60
1	2	5.70
1	3	17.20
1	4	15.85
1	5	14.63
1	6	23.38
1	7	24.14

3.2. Geological Overview of the Study Area

The total area of the study region is approximately 960 km², with a population exceeding 500,000. The geomorphological characteristics of the area exhibit a pronounced stepped pattern, with higher elevations in the central part along the Laoshan axis and lower elevations on both sides. All three major rock types are present in the region. Sedimentary rocks dominate, followed by igneous rocks, and metamorphic rocks occur to a lesser extent.

From the Sinian to the Early Paleozoic Silurian, the region experienced platform-type carbonate sedimentation, forming the first relatively complete sedimentary cover, which regionally and unconformably overlies pre-Sinian formations such as the Zhangbaling Group at different stratigraphic levels. During the Mesozoic, sporadically distributed pyroclastic rocks and intermediate to acidic volcanic lavas were formed. In the Cenozoic Tertiary, a widespread sequence of siltstone to argillaceous siltstone was deposited, forming the second sedimentary cover with a thickness ranging from 50 m to 200 m. This unit is mainly distributed on the terraces of the Chu River and the Yangtze River and constitutes an important foundation for construction sites in the region.

Within the modeling area, the average borehole depth is 61.2 m, and the stratigraphic units encountered are predominantly Quaternary deposits. Overall, the strata can be divided into seven layers, from top to bottom: fill, silt interbedded with fine sand, silty clay, silty clay mixed with coarse sand and gravel, cobble–gravel interbedded with silty clay, weathered residual soil, and bedrock. In this study, these engineering geological units are collectively referred to as “strata” and are numbered from Stratum 1 to Stratum 7.

During the modeling process, geological constraints derived from actual engineering investigation data are incorporated to ensure geological plausibility. For deep regions not penetrated by boreholes, bedrock is adopted as the basal filling unit based on regional geological survey data and stratigraphic contact relationships, thereby ensuring the completeness and reliability of the geological model.

3.3. Experimental Results and Analysis

3.3.1. Performance Evaluation of VA-HDW Feature Matrix Construction

The VA-HDW method relies on the construction of features between stratigraphic spatial points within the input region and individual boreholes. These features are critical to both the classification accuracy and computational efficiency of the VA-HDW method, and the performance of feature matrix construction directly affects the overall effectiveness and efficiency of the proposed approach.

In this experiment, the performance of feature matrix construction was systematically evaluated using a geological borehole dataset from a newly developed urban area. The objective was to assess the computational time required for feature matrix construction under different dataset sizes and upsampling rates. Feature matrix construction involves first analyzing the distance relationships between stratigraphic spatial points and boreholes, computing the feature vector for each spatial point, and then assembling these vectors into a feature matrix. Parallel computing strategies were employed to accelerate this process.

The size of the borehole set, determined by the number of boreholes, directly affects the efficiency of feature matrix construction. When the number of boreholes is fixed, the upsampling rate (defined as the ratio between the number of points after upsampling and the original number of data points) influences the number of stratigraphic spatial points and consequently impacts construction efficiency. Therefore, experiments were conducted to evaluate the time consumption of feature matrix construction under different dataset sizes and upsampling rate settings. For each experimental configuration, the reported results correspond to the average of ten repeated runs. The experimental results are presented in Table 3.

Table 3. Impact of borehole scale and upsampling rate on feature matrix construction time.

Borehole Scale	Upsampling Rate	Sample Point Count	Construction Time (ms)
20	3	384	12.75
	10	1280	25.94
	30	3840	85.07
40	3	804	61.31
	10	2680	123.82
	30	8040	314.91
64	3	1308	125.50
	10	4360	490.53
	30	13,080	1141.98

The experiments were conducted using datasets with borehole scales of 20, 40, and 64. Based on the feature construction method proposed in Chapter 2, feature vectors were constructed for the borehole spatial points obtained after upsampling, and a feature matrix was subsequently formed. The experimental results show that the average computation time increases progressively with both the borehole scale and the upsampling rate. This is because an increase in the number of boreholes leads to a quadratic growth in computational operations, thereby significantly extending the time required for feature construction and model computation. In addition, a higher upsampling rate results in a

linear increase in the number of stratigraphic spatial points, which also contributes to the overall computational cost.

From an empirical perspective, the feature matrices generated using the VA-HDW method achieve the expected performance. Changes in the upsampling rate exhibit a clear linear relationship with feature matrix construction efficiency. An appropriate upsampling rate can improve the quality of the feature matrix; however, excessively high upsampling rates may lead to overfitting. Therefore, in this study, the upsampling rate is set to 10 in order to achieve a balance between feature matrix performance and the risk of overfitting.

3.3.2. Comparative Analysis of VA-HDW and Other Stratum Classification Methods

To verify the superiority of the stratum classification method based on VA-HDW proposed in this paper, a GBARNN neural network was first designed based on the idea of VA-HDW, and the implementation of the VA-HDW method was completed by training this network. Subsequently, based on the classification task, the advantages and disadvantages of the VA-HDW method were evaluated using indicators such as accuracy, F1-score, and the confusion matrix of the classification results. A comparative analysis was conducted between the VA-HDW method and other stratum spatial point classification methods, including KNN, SVM, and GeoPDNN [19]. Among them, the input features of KNN and SVM are both normalized 3D spatial coordinates (x, y, z), while the input features of GeoPDNN refer to the settings in its original study. In addition, the parameters involved in the classification methods themselves have an impact on the results; therefore, a certain degree of parameter tuning was performed on the relevant classifiers, and the experiment only compared the indicators of each method when the optimal accuracy was achieved.

In this experiment, a detailed verification of stratum classification was conducted based on the dataset of a new urban area, aiming to evaluate the comparative effect between the VA-HDW method proposed in this paper and other classification methods.

(1) Parameter Settings and Network Training

In this experiment, the network structure settings of GBARNN and GBDNN are shown in Table 4, including information such as the number of hidden layers, the number of neural units in hidden layers, the loss function, and the optimizer type.

Table 4. Network Structure Settings.

GBDNN	Input Layer	Hidden Layer 1	Hidden Layer 2		Output Layer
	3	6	3		1
GBARNN	Input Layer 64	Hidden Layer 1 256	Hidden Layer 2 256	Hidden Layer 3 64	Output Layer 7

The hyperparameter settings in this experiment are shown in Table 5.

Table 5. Hyperparameter Settings.

Hyperparameter	Value
Training–Validation Set Ratio	8:2
Activation Function	Leaky ReLU (for hidden layers); Softmax (for output layer).
Loss Function	Cross-Entropy Loss Function
Optimizer Type	Adam Optimizer
Learning Rate	0.001
Training Epochs	500
Batch Size	64

The image showing the changes in loss function values and accuracy with the increase in the number of iterations during the training process is presented in Figure 12.

By observing the above figure, it can be seen that after 500 iterations of training, the GBARNN exhibited good convergence during the training process. The loss value decreased while the accuracy increased, and after 500 epochs, the loss value tended to stabilize. This indicates that the network gradually learned the correct stratum classification patterns and converged, demonstrating that the model can effectively learn the features of the stratum classification task and gradually optimize its performance.

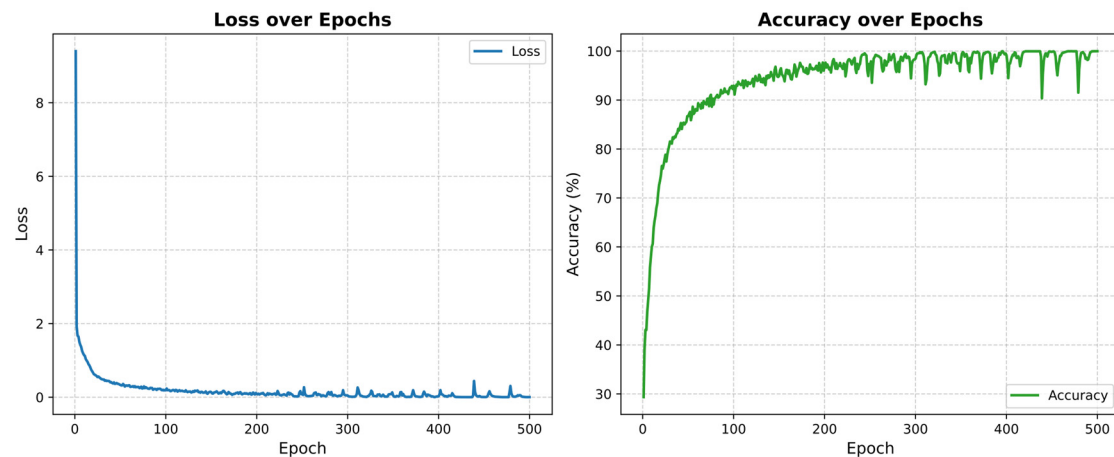


Figure 12. Training Process Diagram of the GBARNN Network.

(2) Analysis of Comparative Experimental Results

In this section, by comparing the GBARNN, KNN, SVM, and GeoPDNN models, their respective performances in the stratum classification task are systematically evaluated. Three indicators, namely classification accuracy, F1-score, and Kappa coefficient, are used for comprehensive assessment. Meanwhile, aiming at the parameter error problem of KNN and SVM, the grid search method is adopted for parameter optimization, and the different parameter ranges and the parameter configuration corresponding to the optimal classification accuracy are recorded in the table. The experimental results are presented in Table 6.

Table 6. Parameter Settings of the Optimized Classification Algorithms.

Classification Algorithm	The Tested Parameter Range	Optimized Parameters
KNN	$n_neighbors \in \{1, 2, 3, \dots, 15\}$	$n_neighbors = 10$
	$leaf_size \in \{20, 30, 40\}$	$leaf_size = 30$
SVM	$p \in \{1, 2\}$	$p = 2$
	$C \in \{0.1, 1, 10, 100\}$	$C = 1$
	$kernel = 'rbf'$	$kernel = 'rbf'$
	$\Gamma \in \{scale, auto\}$	$\Gamma = 'scale'$

From the experimental results in Table 7, it can be observed that in terms of classification accuracy, the traditional machine learning methods KNN and SVM perform relatively poorly in classifying stratum spatial points, while the deep learning models GeoPDNN and the proposed GBARNN show a small gap, and both achieve high classification accuracy. In terms of F1 Score, GeoPDNN and GBARNN still demonstrate strong classification capabilities, showing a good balance between precision and recall. In contrast, the F1 Scores of KNN and SVM remain relatively low, indicating deficiencies in their predictive performance for certain categories.

An in-depth exploration of the distribution of Kappa coefficients reveals that the two deep learning models, GeoPDNN and GBARNN, exhibit extremely high stability. Their Kappa values mostly remain within a high range, leading to the conclusion that after

excluding the interference of random factors, these two methods are more reliable and stable in classification accuracy. In contrast, traditional machine learning algorithms such as KNN and SVM are less reliable, as their Kappa values are much lower and tend to fluctuate significantly when handling multi-category recognition tasks due to external interference.

Table 7. Experimental Results of GBARNN and Other Algorithms.

Indicator	Classification Algorithm	Value(%)
Accuracy	KNN	85.02
	SVM	88.51
	GeoPDNN	91.41
	GBARNN	92.78
F1 Score	KNN	84.15
	SVM	89.27
	GeoPDNN	93.37
	GBARNN	92.42
Kappa	KNN	85.26
	SVM	87.91
	GeoPDNN	89.74
	GBARNN	93.51

Compared with GeoPDNN, GBARNN greatly enhances the deep feature representation capability through the concept of generalized distance between boreholes and stratum spatial points, thereby improving the analytical accuracy of geological attributes of stratum spatial points. This model achieves deep integration of spatial data and borehole information through the generalized distance mechanism, effectively enhancing its adaptability and predictive accuracy in stratum classification tasks. When processing complex stratum data, GBARNN provides more accurate and comprehensive predictive results due to its refined feature extraction characteristics, with slightly improved classification accuracy and Kappa coefficient compared to GeoPDNN. However, despite improvements in feature extraction, GBARNN's over-reliance on certain features may reduce the model's generalization ability, resulting in slightly inferior coordination between precision and recall and a relatively lower overall F1 score.

The experimental results in Figure 13 show that in terms of classification accuracy, the GBARNN neural network outperforms traditional classic algorithms such as KNN, SVM, and GeoPDNN. It also performs the best in terms of the Kappa coefficient. Although there is a gap between GBARNN and the GeoPDNN algorithm in terms of F1 score, GBARNN has obvious advantages in the classification process. To better explore the application value of GBARNN and identify room for optimization, this paper intends to adopt the method of independent sample testing, select specific target strata for classification effect research, and use confusion matrices to analyze the recognition effects and improvement space of different models under different stratum attributes. Subsequently, the classification performance of the four algorithms on a specific stratum will be studied.

Figure 14 presents the stratum classification results based on the KNN algorithm, with a comprehensive evaluation using a confusion matrix. Statistics show that the classification accuracy of this method for each stratum ranges from 80% to 91%, indicating a certain level of recognition capability. However, although the overall performance is relatively stable, obvious misjudgments occur in the 5th and 6th strata. This suggests that due to its over-reliance on the assumption of local similarity, the KNN model struggles to accurately capture the nonlinear relationships between strata when dealing with complex geological structures.

As shown in Figure 15, the average accuracy of the SVM algorithm in classifying various geological strata ranges from 86% to 91%, indicating that the SVM algorithm has

strong stability. Compared with the KNN algorithm, the SVM algorithm is more reliable in classification performance, with a smaller fluctuation range in prediction results and stronger anti-interference ability, which suggests that the SVM algorithm can more accurately identify differences between samples in geological stratum classification tasks. Although the SVM algorithm has stronger classification consistency than the KNN algorithm, it still cannot avoid misjudgments for some datasets with large category differences. This may be related to the insufficient ability of the SVM algorithm to explain the internal complexity of some geological stratum data.

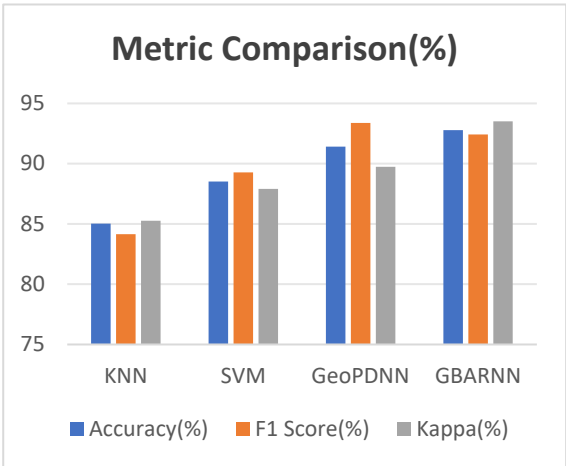


Figure 13. Histogram of Metrics for Four Stratum Classification Methods.

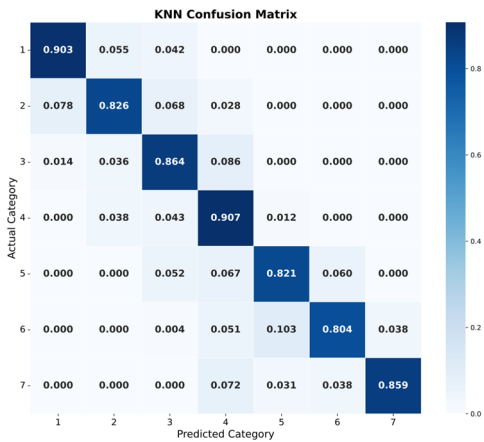


Figure 14. KNN Confusion Matrix.

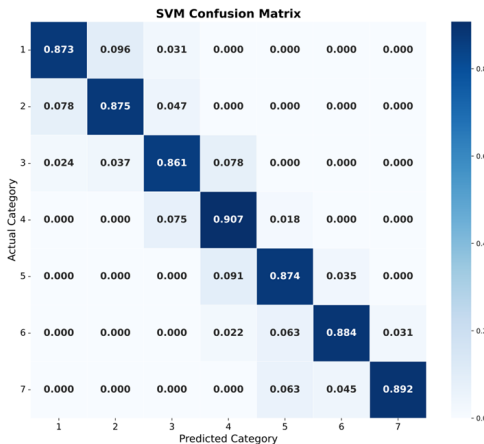


Figure 15. SVM Confusion Matrix.

Figure 16 shows the confusion matrix of the GeoPDNN algorithm. According to the classification performance evaluation results of the GeoPDNN model, its accuracy for samples of various strata is mostly between 89% and 93%, indicating that the GeoPDNN algorithm has obvious advantages in stratum classification and excellent ability in balanced prediction across categories. Compared with traditional methods, the GeoPDNN algorithm achieves higher overall accuracy in multi-stratum recognition, with a more balanced error distribution. This demonstrates that the GeoPDNN algorithm has stronger generalization ability and adaptability to complex geological environments, thereby reducing the risk of misjudgment.

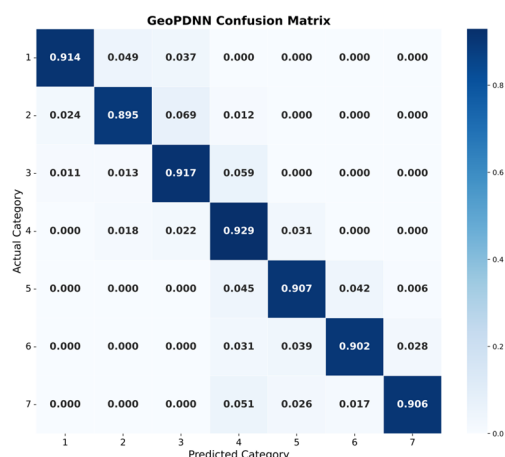


Figure 16. GeoPDNN Confusion Matrix.

As shown in Figure 17, the analysis of the confusion matrix constructed using the GBARNN algorithm indicates that it has exhibited good performance in various rock stratum classification tasks, with the overall classification accuracy basically stable between 91% and 94%, which suggests that this algorithm has advantages in stratum classification and can maintain high recognition accuracy in complex geological data environments; in terms of the distribution of classification errors between different rock stratum sequences, the fluctuation exhibited by GBARNN is relatively low, which further highlights its generalization ability and stability characteristics in multi-source stratum samples.

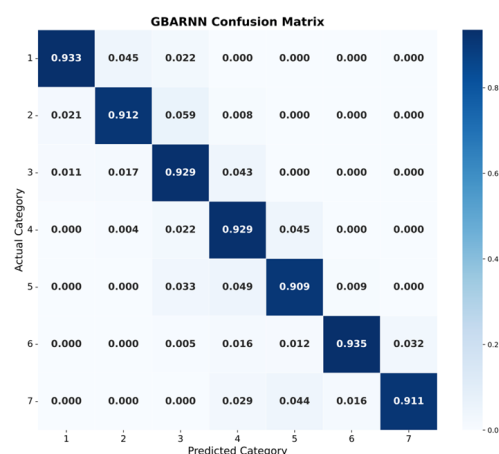


Figure 17. GBARNN Confusion Matrix.

In special scenarios where stratum classification accuracy is relatively low, the GBARNN algorithm exhibits obvious stability, achieving a classification accuracy of 91.2%. This well demonstrates the model's robustness in complex category determination and its

efficient feature representation capability. By adopting the generalized distance measurement method and improving the feature formation mechanism, GBARNN can accurately identify many core features containing abundant spatial distribution information, thereby improving the overall classification performance.

Compared with traditional machine learning algorithms, GBARNN has significantly improved classification accuracy; its optimized design scheme has greatly reduced the misclassification probability. When dealing with complex correlation structures and high-dimensional features in geological data, it exhibits excellent robustness and anti-interference ability. In contrast to other classic methods (KNN, SVM), GBARNN, relying on the deep learning framework, can more accurately mine the nonlinear features in the data, thereby improving the classification effect.

3.3.3. Geological Body Modeling Experiments Based on GBARNN Classification Results

This subsection mainly conducts geological body modeling experiments based on the classification results of the GBARNN model. The specific procedures are as follows:

1. The stratigraphic information of the study area is analyzed to train a GBARNN model for stratum classification.
2. A spatial bounding box covering the study area is defined, within which several detection boreholes are arranged. Stratigraphic sampling points to be classified are generated along these boreholes at fixed intervals through resampling.
3. The trained classification model is applied to classify the stratigraphic attributes of the sampling points along each detection borehole.
4. The boundary point coordinates (X , Y , Z) of each stratigraphic surface are aggregated, and kriging interpolation is employed to fit constrained stratigraphic surfaces. The volume between two adjacent stratigraphic surfaces is regarded as a single stratigraphic unit.
5. The spatial bounding box is discretized into a set of three-dimensional grids. Based on the marching cubes algorithm, the surface models are visualized as isosurfaces. Subsequently, stratigraphic attributes are assigned to all grid cells according to the spatial relationships between the grid cells and the isosurfaces, thereby constructing the geological body model.

To better represent geological body boundaries, the outermost boreholes are extracted as boundary constraints to ensure that the constructed model more closely matches real geological conditions. Figure 18 illustrates the bounding box constrained by geological body boundaries.

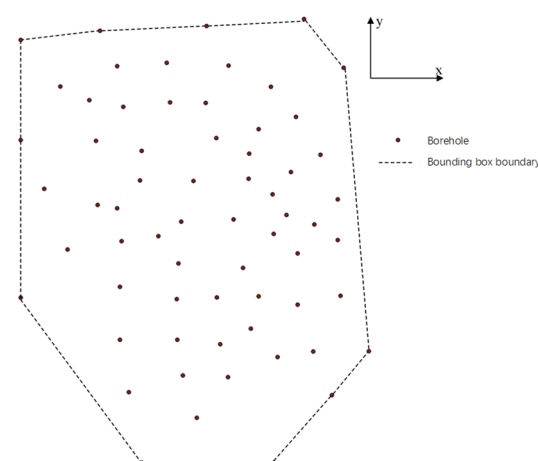


Figure 18. Hatching distribution map.

To ensure the spatial accuracy of the model, the grid resolution is set to $256 \times 256 \times 256$. Within the bounding box of the target area, 500 detection boreholes are uniformly arranged, and each detection borehole is discretized into 256 sampling intervals along the depth direction to accurately capture stratigraphic variations. Based on these settings, the modeling workflow for a single stratigraphic unit is illustrated in Figure 19.

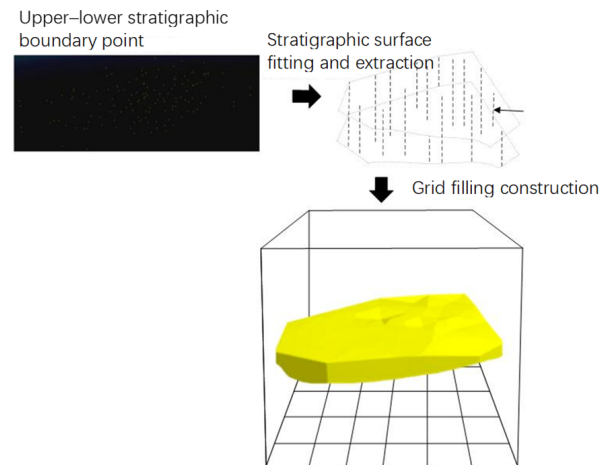


Figure 19. Formation process of the silty clay mixed with coarse sand and gravel layer.

Based on the above procedures, a three-dimensional geological body model of a newly developed urban area is constructed, and smoothing processing is applied to the model. The resulting models are shown in Figures 20 and 21.

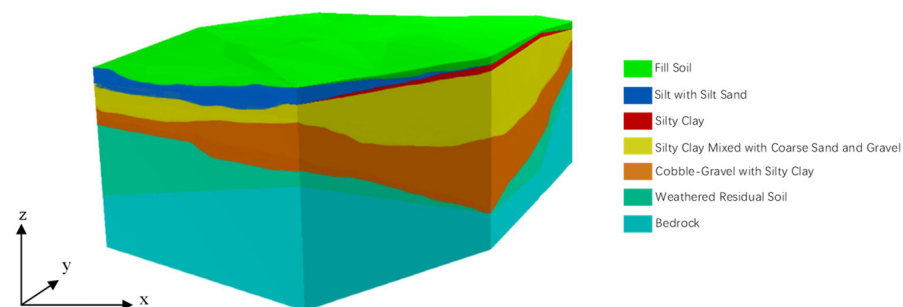


Figure 20. Geological body model.

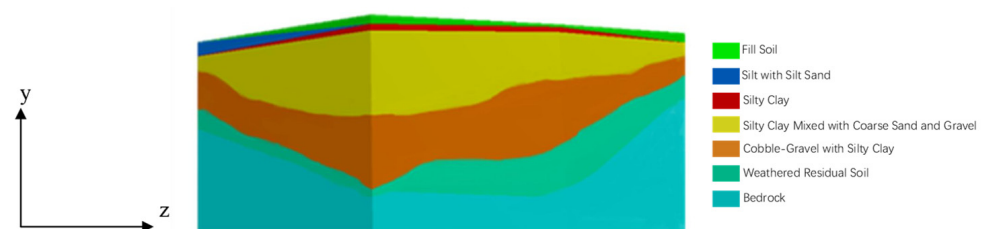


Figure 21. Side view of the geological body model.

To evaluate the consistency between the constructed geological body and the actual borehole data, a comparative analysis is conducted between the geological cross-sections extracted from the model and the real borehole information. This analysis focuses on whether the stratigraphic distributions along the cross-sections are consistent with the strata revealed by the boreholes, as well as whether key characteristics such as stratigraphic thickness are in agreement. Through this comparison, the accuracy and reliability of the geological body model in representing real geological conditions can be assessed.

As shown in Figure 22, two cross-sections, S1 and S2, are established within the study area to validate the accuracy of the proposed modeling method. Figure 23 further presents comparisons between the S1 and S2 cross-sections and the surrounding actual borehole data. Since the geological body model is constructed based on detection (virtual) boreholes rather than being directly fitted to the real boreholes, minor discrepancies with the actual borehole information are observed. Nevertheless, from an overall perspective, the modeling results are generally consistent with the actual geological characteristics.

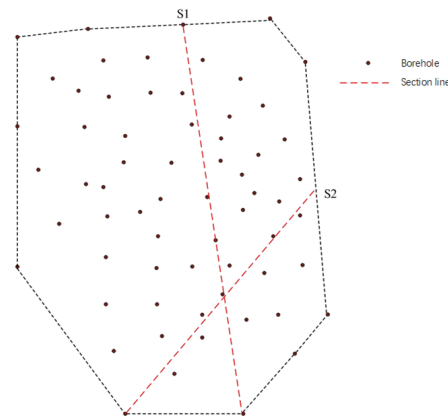


Figure 22. Cross-section line distribution map.

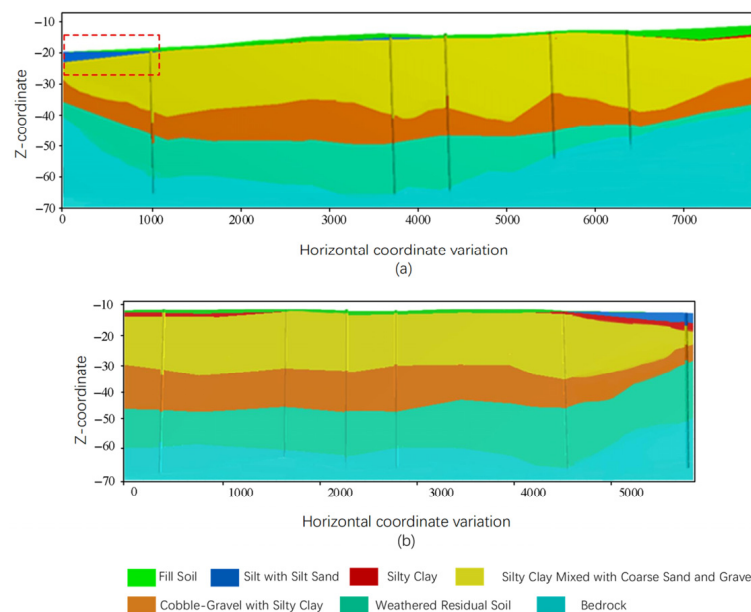


Figure 23. Comparison between cross-sections S1 and S2 and nearby actual boreholes, (a) cross-section S1; (b) cross-section S2.

The proposed modeling approach integrates the VA-HDW method and therefore possesses a certain capability for characterizing stratigraphic pinch-out phenomena. To evaluate the overall performance of the method in modeling pinch-out strata, the pinch-out feature highlighted by the red box on cross-section S1 is selected for detailed analysis. Since the present study focuses only on the first five stratigraphic units, the remaining portions of the corresponding boreholes are masked in the analysis. As shown in Figure 24, the selected area exhibits a high potential for pinch-out occurrence. The modeling results along cross-section S1 are consistent with this observation, thereby confirming the presence of pinch-out in the constructed geological model. These results indicate that the proposed modeling method demonstrates a certain capability in modeling stratigraphic pinch-out features.

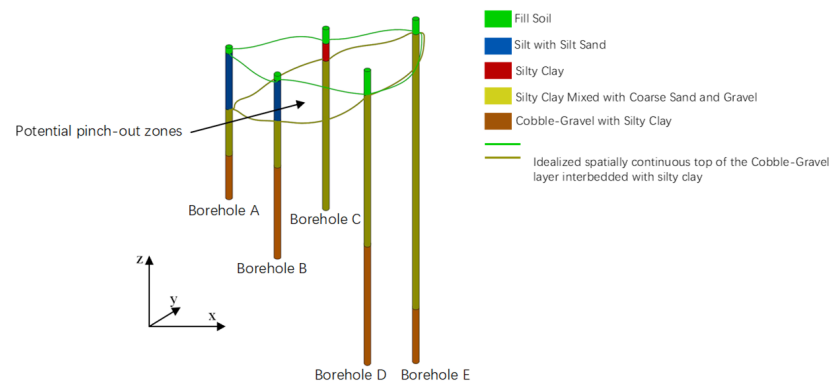


Figure 24. Pinch-out analysis.

In this study, the geological body model is constructed based on detection (virtual) boreholes, while the actual boreholes are used only as input training data. As a result, the spatial coupling between detection boreholes and actual boreholes is relatively weak. Therefore, analyzing the relationship between the average stratigraphic thickness derived from the stratigraphic surfaces generated by detection boreholes and the average stratigraphic thickness reflected by actual boreholes provides an objective means to evaluate the performance of the proposed modeling method. In this study, a borehole fitting degree index is adopted to quantitatively assess the fitting accuracy of stratigraphic thickness. This index effectively reflects the capability of the model to represent geological structures. The calculation formula is given as follows:

$$\delta_i = 1 - \frac{|H_{model}^i - H_{real}^i|}{H_{real}^i} \quad (17)$$

In the above formulation, i denotes the stratigraphic layer index, and δ_i represents the borehole fitting degree of the i stratum. H_{model}^i denotes the average thickness of the i stratum at the actual borehole locations derived from the geological body generated by the classification model, while H_{real}^i denotes the corresponding actual average stratigraphic thickness obtained from real borehole data. A smaller value of δ_i indicates a larger discrepancy between the model and the real geological conditions, whereas a value closer to 1 implies better agreement between the modeling results and the actual geological situation. In this study, the borehole fitting degree of the VA-HDW-based method is calculated, and the fitting degree results for each stratigraphic layer are summarized in Table 8.

Table 8. Results of borehole fitting degree calculation.

Stratum	VA-HDW
1	0.9179
2	0.9032
3	0.8763
4	0.9114
5	0.8476
6	0.9157
7	0.9181
Mean value	0.8986

4. Discussion

This paper proposes a stratigraphic classification method based on vertical alignment and horizontal distance weighting for stratigraphic spatial point classification. This approach fully exploits geological features, enhances the classification accuracy of strati-

graphic spatial points, and provides a solid foundation for geological body modeling. Although the aforementioned research has achieved some phased successes, many areas still require in-depth study and exploration:

1. Geological boreholes are not necessarily vertical. For instance, deviated wells are common in hydrocarbon reservoir exploration. To extend the VA-HDW method to deviated well scenarios, additional well trajectory information—such as well inclination angle and azimuth—must be incorporated. This requires transforming the original borehole coordinates to establish a consistent three-dimensional spatial reference system. Building upon this, integrating well trajectory correction with the proposed method and conducting experimental validation on inclined well datasets will serve as a key research direction for future work.
2. This paper employs cross-section analysis and borehole fit metrics to compare modeling results. Additional metrics should be considered to evaluate modeling outcomes from diverse perspectives. For instance, establishing geological probability models for each 3D geological body model or utilizing information entropy could assess each 3D geological model from an uncertainty standpoint, thereby validating the feasibility of the modeling approach.

5. Conclusions

This paper focuses on 3D geological modeling methods for borehole data; aiming at issues such as low classification accuracy and difficulties in identifying geological data features in machine learning-based 3D geological modeling of borehole data, it studies the classification methods of stratum spatial points, and by constructing key geological features and measuring the influence degree between stratum spatial points and boreholes, the classification effect of stratum spatial points is improved. This paper proposes a stratum classification method based on VA-HDW, which constructs features between stratum spatial points and boreholes, contributing to the effective expression of geological features, and based on this method, a neural network structure is designed; through training this network, the accuracy and reliability of stratum classification are enhanced, and the quality of 3D geological body modeling is improved. Experimental verification shows that the method proposed in this paper can enhance the classification effect of stratum spatial points and improve the quality of 3D geological body modeling.

Author Contributions: Z.H. wrote the test codes and performed the experiments. H.L. designed a stratum classification method based on VA-HDW and, building upon this, designed a dedicated neural network model: GBARNN. C.Z. analyzed the experimental dataset and revised the manuscript. H.L. wrote the paper, and all authors reviewed the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Key Research and Development Program of China (2022YFF0801203, 2022YFF0801200) and the National Natural Science Foundation of China (41972306).

Data Availability Statement: The data presented in this study are available on request from the corresponding author due to privacy and legal restrictions.

Acknowledgments: Thanks for the reviewers' constructive comments and they really help us improve this paper.

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

References

1. Zhang, Z.Q.; Wang, G.W.; Liu, C.; Cheng, L.Z.; Sha, D.M. Bagging-based positive-unlabeled learning algorithm with Bayesian hyperparameter optimization for three-dimensional mineral potential mapping. *Comput. Geosci.* **2021**, *154*, 104817. [\[CrossRef\]](#)
2. Thibaut, R.; Laloy, E.; Hermans, T. A new framework for experimental design using Bayesian Evidential Learning: The case of wellhead protection area. *J. Hydrol.* **2021**, *603*, 126903. [\[CrossRef\]](#)
3. Hoyer, A.S.; Klint, K.E.S.; Fiandaca, G.; Maurya, P.K.; Christiansen, A.V.; Balbarini, N.; Bjerg, P.L.; Hansen, T.B.; Moller, I. Development of a high-resolution 3D geological model for landfill leachate risk assessment. *Eng. Geol.* **2019**, *249*, 45–59. [\[CrossRef\]](#)
4. Wang, S.; Wang, W.; Sun, D.; Ma, Z. 3D Modeling Method of Geological Faults Based on Section Map. *J. Geomat.* **2016**, *41*, 59–63. [\[CrossRef\]](#)
5. Wu, Z.C.; Guo, F.S.; Jiang, Y.B.; Luo, J.Q.; Hou, M.Q. Methods of Three-dimension Geological Modeling Based on Geological Sections. *Geol. Explor.* **2016**, *52*, 363–375.
6. 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*, 103390. [\[CrossRef\]](#)
7. Ming, J.; Pan, M. Quick borehole data deciphering based on stratum tagging. *Chin. J. Geotech. Eng.* **2009**, *31*, 692–698.
8. Tang, B.Y.; Wu, C.L.; Li, X.C.; Chen, Q.Y.; Mu, H.T. A fast progressive 3D geological modeling method based on borehole data. *Rock Soil Mech.* **2015**, *36*, 3633–3638.
9. Guo, J.T.; Zheng, Y.F.; Liu, Z.B.; Wang, X.L.; Zhang, J.Q.; Zhang, X.Z. Pattern-Based Multiple-point Geostatistics for 3D Automatic Geological Modeling of Borehole Data. *Nat. Resour. Res.* **2025**, *34*, 149–169. [\[CrossRef\]](#)
10. Guo, J.T.; Liu, Y.H.; Han, Y.F.; Wang, X.L. Implicit 3D Geological Modeling Method for Borehole Data Based on Machine Learning. *J. Northeast. Univ. (Nat. Sci.)* **2019**, *40*, 1337–1342. [\[CrossRef\]](#)
11. Xiong, J.Q.; Liu, X. A 3D Geological Model of the North One Mining Area of Gubei Coal Mine Based on the Support Vector Machine. *Sci. Technol. Eng.* **2022**, *22*, 8194–8199.
12. Zhu, J.S.; Wang, S.; Bai, J.; Xu, Z.X.; Chen, M.H.; Li, Z.Q.; Liu, X.; Zhang, Z.H.; Liu, X.Y. An Improved KNN Algorithm Method Using Finite Boreholes for Predicting Full-area Geological Features. *Tunn. Constr.* **2023**, *32*, 348–358.
13. Ji, X.J.; Lu, X.Y.; Guo, C.H.; Pei, W.W.; Xu, H. Predictions of Geological Interface Using Relevant Vector Machine with Borehole Data. *Sustainability* **2022**, *14*, 10122. [\[CrossRef\]](#)
14. Titos, M.; Bueno, A.; García, L.; Benítez, C. A Deep Neural Networks Approach to Automatic Recognition Systems for Volcano-Seismic Events. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* **2018**, *11*, 1533–1544. [\[CrossRef\]](#)
15. Zhang, T.F.; Tilke, P.; Dupont, E.; Zhu, L.C.; Liang, L.; Bailey, W. Generating geologically realistic 3D reservoir facies models using deep learning of sedimentary architecture with generative adversarial networks. *Pet. Sci.* **2019**, *16*, 541–549. [\[CrossRef\]](#)
16. Lyu, B.; Wang, Y.; Shi, C. Multi-scale generative adversarial networks (GAN) for generation of three-dimensional subsurface geological models from limited boreholes and prior geological knowledge. *Comput. Geotech.* **2024**, *170*, 106336. [\[CrossRef\]](#)
17. Lyu, B.R.; Wang, Y.; Miao, C.; Yao, J.C.; Shum, L.K.W.; Wong, A.L.; Ho, R.C.M. Fusion of Limited Site-Specific Borehole Logs and Geophysical Data from a Different Site for Three-Dimensional Subsurface Geological Modeling Using Multiscale Generative Adversarial Network. *J. Geotech. Geoenviron. Eng.* **2025**, *151*, 04025087. [\[CrossRef\]](#)
18. Hang, Z.Q.; Xue, T.; Chen, J.P.; Shi, Y.J.; Yin, Z.H.; Cui, Z.J.; Zhou, G.Y. A 3D Geological Modeling Method Using the Transformer Model: A Solution for Sparse Borehole Data. *Minerals* **2025**, *15*, 301. [\[CrossRef\]](#)
19. Guo, J.T.; Xu, X.C.; Wang, L.Y.; Wang, X.L.; Wu, L.X.; Jessell, M.; Ogarko, V.; Liu, Z.B.; Zheng, Y.F. GeoPDNN 1.0: A semi-supervised deep learning neural network using pseudo-labels for three-dimensional shallow strata modelling and uncertainty analysis in urban areas from borehole data. *Geosci. Model Dev.* **2024**, *17*, 957–973. [\[CrossRef\]](#)
20. Gonçalves, I.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. [\[CrossRef\]](#)
21. He, Z.X.; Xu, X.L.; Peng, P.G.; Wang, L.G.; Tian, S.C. A deep learning-driven three-dimensional geological modeling method using sparse borehole sampling data. *Measurement* **2025**, *256*, 118461. [\[CrossRef\]](#)

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