

## Research Article

# Research on Stratum Identification Method Based on TBM Tunneling Characteristic Parameters

Wei Wu <sup>1,2</sup>, Jingbo Guo <sup>1,2</sup>, Jie Li,<sup>1,2</sup> Ji Sun,<sup>1,2</sup> Haoran Qi,<sup>1,2</sup> and Ximing Chen<sup>1,2</sup>

<sup>1</sup>School of Mechanical Engineering, Shijiazhuang Tiedao University, Shijiazhuang 050043, China

<sup>2</sup>State Key Laboratory of Mechanical Behavior and System Safety of Traffic Engineering Structures, School of Mechanical Engineering (Shijiazhuang Tiedao University), Hebei, Shijiazhuang 050043, China

Correspondence should be addressed to Jingbo Guo; [guojingbo66@163.com](mailto:guojingbo66@163.com)

Received 10 May 2022; Accepted 12 October 2022; Published 29 October 2022

Academic Editor: Daniele Salvati

Copyright © 2022 Wei Wu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In order to obtain continuous stratum information during TBM tunneling, using TBM tunneling parameters, stratum recognition is carried out through the K-nearest neighbor model, and the model is improved by the entropy weight method to improve the stratum recognition rate. By analyzing the correlation between TBM tunneling characteristic parameters and stratum, the tunneling characteristic parameter vector which is most sensitive to the stratum is obtained by sensitivity analysis, and the stratum recognition model based on the K-nearest neighbor algorithm is established. Aiming at the problem that the model has a large error in complex formation recognition, a formation recognition model based on the entropy weight K-nearest neighbor algorithm is established, and the wrong data of the K-nearest neighbor model is recalculated. The recognition rate of the stratum in the new model is increased from 90.95% to 98.55%. The results show that the K-nearest neighbor model has a better recognition effect for the interval with single stratum distribution, and the recognition rate of entropy weight K-nearest neighbor model for complex stratum is significantly improved, which provides an effective method to obtain stratum information by using tunneling characteristic parameters.

## 1. Introduction

With the increasing demand for tunnel construction and underground engineering in China, TBM is widely used as special tunneling equipment. In the early stage of TBM selection and construction, the geological survey of the whole construction section is required. Due to China's vast territory and complex geological conditions, the TBM construction process often faces complex geology such as a boulder, upper soft, and lower hard. Considering the construction cost and survey methods (usually discontinuous drilling sampling), accurate geological characteristics cannot be obtained from geological survey data. In the process of tunnel construction, the objective and complex geological conditions are one of the main reasons restricting the safe and efficient construction of tunnels, which may lead to the damage of the TBM, affect the construction period, and more seriously endanger the safety of tunnel construction

[1–4]. At present, the advanced geological prediction methods are mainly divided into a geological method, a physical exploration method, and a drilling method [5–7]. However, considering the large differences in the scope of its use and sensing objects, each has its advantages and disadvantages, and a single sensing method is difficult to be effectively guaranteed in terms of accuracy. Therefore, it is particularly important to accurately identify geological characteristics during TBM tunneling.

As the external performance of the TBM perception of geology, studying the relationship between tunneling parameters and geological conditions is conducive to the recognition of geological characteristics. Relevant scholars at home and abroad have conducted extensive research on the correlation between TBM tunneling parameters, derivative parameters of main tunneling parameters, and strata. Zhao et al. [8] analyzed the correlation between the tunneling parameters of TBMs and various strata in different projects

and found that the average tunneling speed and the average penetration were quite different. The thrust and torque fluctuated significantly higher in hard rock strata than in other strata, and the rest parameters changed little. Ghasemi et al. [9] and Delisio and Zhao [10] respectively used the fuzzy logic theory method, support vector machine, and multiple regression analysis methods to establish a regression model to predict TBM tunneling speed under hard rock conditions and predict TBM penetration. Guo et al. [11] proposed a three-stage method to predict the collapse position of a TBM tunnel by predicting torque and thrust based on the data of noncollapse area from the construction of a long-short-term memory model. Therefore, it is feasible to invert stratum information through shield tunneling parameters.

Some scholars use tunneling parameters to classify rocks or identify certain geologies. Elbaz et al. [12] involved the expression and classification of geological information in the research of optimizing the performance of the cutter head driver, adopted a coding method to define the correction index of the physical meaning of the input parameters to quantify the geological characteristics, and combined with the mechanical properties of the soil and other characteristics to divide the soil layers into three categories, which improved the accuracy of geological retrieval. Based on the tunneling parameters, Liu et al. [13] proposed the correction of specific energy and constructed the recognition model of boulder geology. Using BP neural network technology, the recognition model was established to realize the recognition of solitary rock geology. Huang and She [14] classified the surrounding rock grade based on the geological records of slag and TBM and established the extension evaluation method of surrounding rock stability classification based on extension theory. Hou et al. [15] proposed a superimposed ensemble classifier for real-time prediction of rock mass classification using TBM operation data, in which the hyper-parameters of each classifier were optimized by the grid search method. Compared with a single classifier, the superimposed ensemble classifier has better performance, and has stronger learning ability for small samples and unbalanced samples. Liu et al. [16] proposed an ensemble learning model based on a classification regression tree and AdaBoost algorithm to predict surrounding rock classification and integrated a few over-sampling techniques to solve the imbalance problem of rock classification in the database. Yan et al. [4, 17] realized the prediction of geological characteristics by integrating grid search (GS) and K-fold cross-validation (K-CV) into the overlay classification algorithm (SCA) according to the data generated by shield tunneling and borehole data. Therefore, in addition to the identification of a single stratum, it is very necessary to identify the combined stratum of the tunnel section.

This paper is based on the Guangfo Ring Line Shield Tunnel Project in Guangdong Province. In the section where the moderately and completely weathered stratum contains boulders, the muck is viscous and difficult to flow, and it is easy to form a hard “mud cake” in the cutter head and the chamber (as shown in Figure 1), which aggravates the tool loss. The torque of the cutter head and the thrust resistance



FIGURE 1: “Mud cake” at cutter head opening.

of the shield increases rapidly. The auger cannot be excavated normally, and the shield cannot be pushed normally. The tunnel project adopts the construction method of a combination of open-cut method and TBM method, in which the surrounding rock of the shallow section of the tunnel is classified as Grade V~VI, and the surrounding rock of the deep section where the rock mass is complete is classified as Grade III~IV. The weathering of the granite section is uneven, the bedrock surface undulates greatly, and there are boulders distributed locally on the top and body of the tunnel. It is difficult to identify the composite strata by the previous methods.

In the process of principal component analysis of shield tunneling parameters, the stratum is introduced, and the tunneling parameters that can best reflect the stratum are extracted. Through sensitivity analysis, the combination of tunneling parameters that are most sensitive to stratum identification through the model is selected, which reduces the impact of redundant features on the identification rate, reduces the dimension of characteristic parameters, and has certain innovations in the method of feature selection. In this paper, the EWM-KNN model based on the entropy weight method (EWM) and KNN model is established for composite strata. The problem of identifying strata by shield tunneling parameters is solved. Especially for the project studied in this paper, the identification of the section containing boulders in the completely weathered stratum has a recognition rate of 98.55%, which is about 8% higher than that of the KNN model. This model has certain innovations in the identification of composite strata.

## 2. Characteristic Parameters Processing and Correlation Analysis

Through the obtained actual TBM tunnel engineering data, the different geotechnical combinations in the tunnel line are analyzed. Combined with the characteristic value of the foundation bearing capacity of rock and soil, uniaxial saturated compressive strength and surrounding rock grade, the different rock and soil combinations are divided, and the geological code is rewritten. The tunneling data during tunnel construction are processed to analyze the correlation between tunneling data and geological code. Based on a tunnel project in Guangdong Province, an earth pressure balance shield (EPB) machine is used, and the open-cut

method combined with the shield method are adopted in the project. The total length of the upward line tunnel is 3100.19 m, of which the EPB tunneling is 982.80 m and the diameter of the cutter head excavation is 9.15 m (the main design parameters of the shield machine are shown in Table 1). This section of the tunnel mainly passes through residual soil, fully weathered granite, strongly weathered granite, and moderately weathered granite. It contains two rock sections where boulders may occur.

*2.1. Establishment of Stratigraphic Code.* According to the geological survey data, the geological survey map can be obtained, as shown in Figure 2. In combination with the national standard Code for Geotechnical Engineering Investigation of Urban Rail Transit (GB 50307-2012) [18] and Code for Design of Building Foundation (GB 50007-2011) [19], according to the characteristics of each rock and soil layer and engineering geological conditions, the surrounding rock types of rock and soil layers of TBM tunnel construction in this project are shown in Table 2.

It can be seen from Table 2 that the residual soil and fully weathered granite belong to grade VI and grade V respectively in the classification of surrounding rock, but the characteristic values of foundation bearing capacity are very similar, and the structural characteristics and complete state are relatively loose. The actual construction process is reflected in the TBM tunneling data, so the residual soil, fully weathered granite, and the combined strata of the two are marked as code No. 1. According to the different combinations of rock and soil layers in the TBM tunneling section, the strata and their combination marks are shown in Table 3.

*2.2. Processing of Tunneling Parameters.* In the study of TBM tunneling performance, parameters such as total thrust, cutter head torque, tunneling speed, cutter head rotational speed, and average soil pressure are usually selected for analysis. Taking the total thrust as an example, the sampling period of the time-domain signal in the TBM tunnel construction project is 1 min, that is, a group of data is collected every 60 s, as shown in Figure 3(a). Considering the interference of the data in the start-stop stage and the data in the nonexcavation process, in this paper, the invalid data in the continuously collected data are removed, and the valid data are spliced to obtain the test data. The binary discriminant method is that in a group of data collected at a certain time, as long as the value of one of the tunneling parameters is zero, all data at that time will be deleted to obtain various tunneling data under the normal tunneling state [13], as shown in Figure 3(b).

The data collected in practical engineering applications often have a lot of noise, and the signal jitter is serious. Therefore, by smoothing the outliers and noise data, the trend of various tunneling data over time is obtained, as shown in Figure 3(c).

The comparison between the original data and the processed data is shown in Figure 3. From the results before and after the total thrust data processing, it is more continuous and intuitive to reflect the variation of tunneling

TABLE 1: Main performance parameters of EPB.

Name	Parameter
Excavation diameter (mm)	9150
The opening rate of cutter head (%)	35
Main drive speed (rpm)	0~2.8
Main drive power (kW)	12 × 250
Rated torque (kN·m)	17960
The thrust of the propulsion cylinder (kN)	70000
Rated torque of screw conveyor (kN·m)	249

parameters with geology in the tunneling process, which provides a data basis for the correlation analysis between tunneling parameters and strata.

*2.3. Correlation Analysis between Tunneling Parameters and Strata.* The main parameters of TBM tunneling include total thrust, cutter head torque, tunneling speed, penetration, and average earth pressure. The thrust factor and tunneling-specific energy derived from the main tunneling parameters can reflect the tunneling performance of the shield machine by integrating the characteristics of various parameters [20].

In the process of TBM tunneling, the penetration is affected by the tunneling speed and the rotational speed of the cutter head, which is defined as the distance of each rotation of the cutter head, and can directly reflect the tunneling efficiency. With other tunneling parameters unchanged, the penetration decreases with the increase of surrounding rock strength. Under the same formation conditions, the penetration increases with the increase of the total thrust, so the thrust factor is introduced to represent the total thrust required by the unit penetration, reflecting the boreability of the formation. The thrust factor  $F'$  is defined as follows [21], and its time domain is shown in Figure 4.

$$F' = \frac{F}{p}, \quad (1)$$

where  $F$  is the total thrust (kN),  $p$  is the penetration (mm/r), which represents the forward driving distance of TBM when the cutter head rotates for one circle.

From the perspective of energy, when other tunneling parameters remain unchanged, the increase of cutter head torque, cutter head speed, or total thrust will consume more tunneling energy. By calculating the energy consumption per unit volume of rock and soil to characterize the boreability of rock and soil, the specific energy (SE) of TBM tunneling has a strong correlation with rock strength [22], which is defined as follows, and the time domain diagram is shown in Figure 5.

$$SE = \frac{T\omega + Fv}{\pi R^2 v}, \quad (2)$$

where  $T$  is the cutter head torque (kN·m),  $\omega$  is the cutter head rotational speed (r/min),  $v$  is the tunneling speed (mm/min), and  $R$  is the cutter excavation radius (m).

Pearson correlation principle is used to analyze the correlation between tunneling parameters, derivative

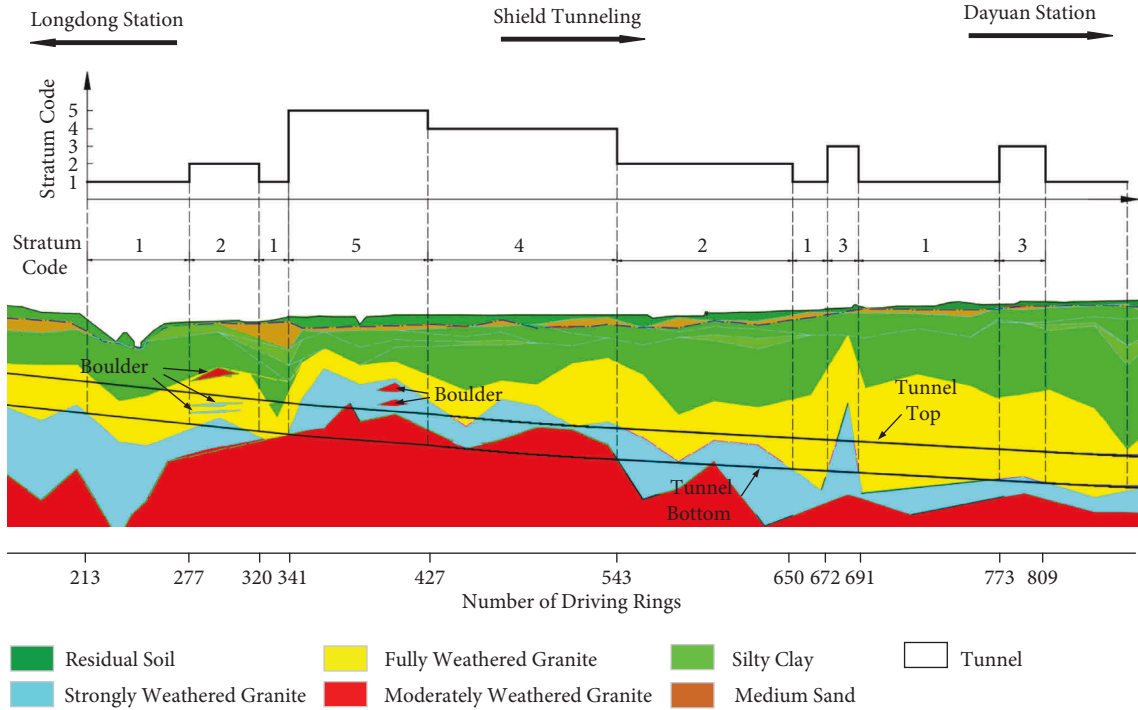


FIGURE 2: Stratum distribution and corresponding stratum code.

TABLE 2: Rock and soil layer characteristics, construction conditions, and surrounding rock classification\*.

Stratum	Structural characteristics and complete state	Characteristic value of foundation bearing capacity (kPa)	Classification of surrounding rocks	Classification of earth and stone
Residual soil	Loose or soft	180	VI	Ordinary soil
Fully weathered granite	Loose structure	200	V	Hard soil
Strongly weathered granite	Inlaid structure of stone and gravel	500	IV	Soft rock
Medium weathered granite	Large block masonry structure, local block stone, gravel mosaic structure	1500	III	Hard stone
Lonely stone	Inlaid structure of stone and gravel	1500	III	Hard stone
Fully weathered granite	Inlaid structure of stone and gravel	500	IV	Soft rock

\*Boulders can be considered similar to moderately weathered bedrock.

TABLE 3: Stratum section combination and corresponding stratum code.

Stratum section combination	Stratum code
Residual soil, fully weathered granite, residual soil, and fully weathered granite	1
Fully weathered granite and strongly weathered granite	2
Strongly weathered granite	3
Strongly weathered granite and moderately weathered granite	4
Boulder	5

parameters, and strata [13], as shown in the following equation, and the correlation results are shown in Table 4.

$$\rho(X, Y) = \frac{E[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y}, \quad (3)$$

where  $X$  and  $Y$  represent two parameters for correlation analysis, respectively.

The correlation coefficient  $\rho$  is a real number between  $(-1, 1)$ . When  $\rho \in [-1, 0]$ , there is a negative correlation between variables. When  $\rho \in (0, 1)$ , there is a positive

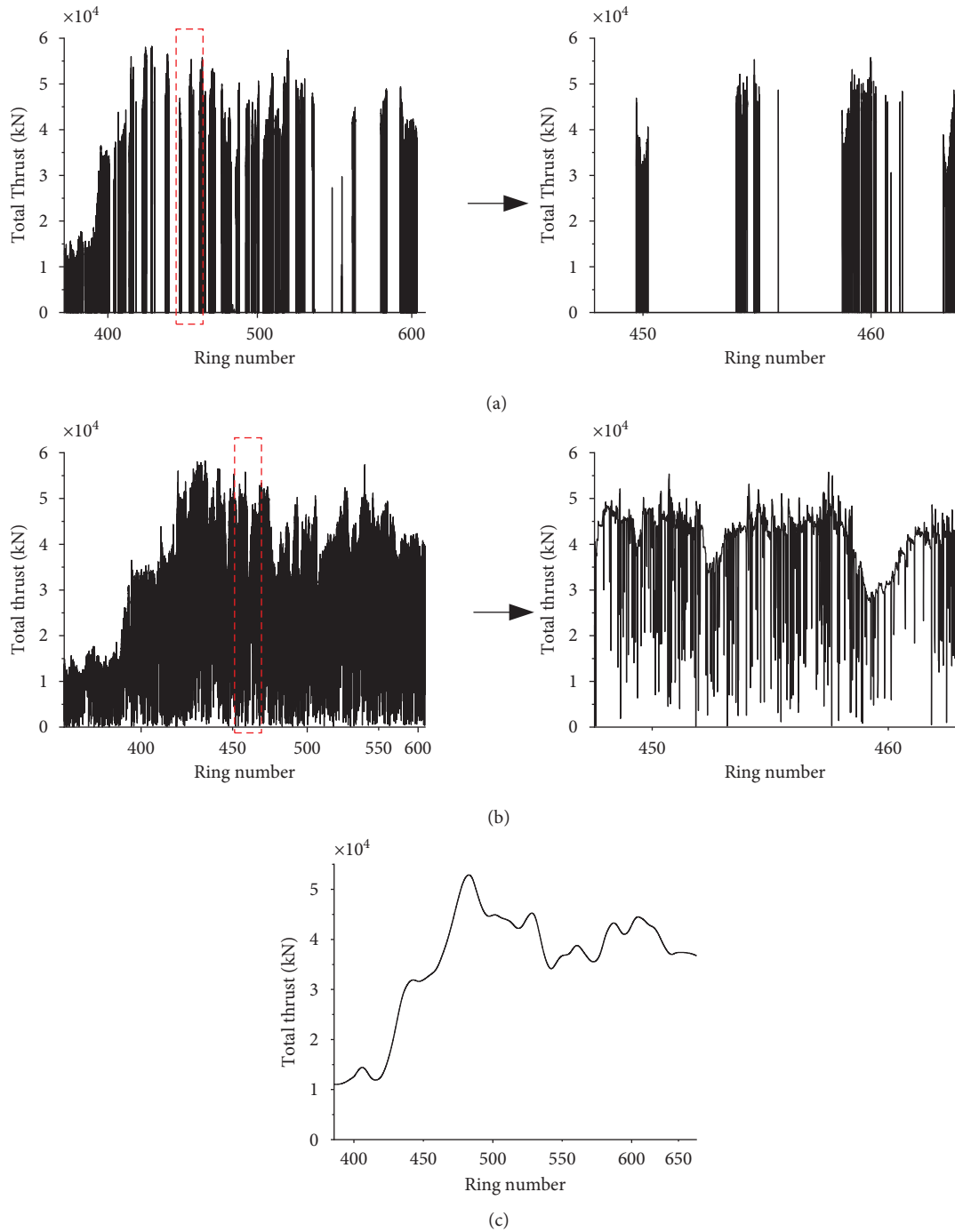


FIGURE 3: Comparison chart of total thrust before and after data processing: (a) original total thrust and partially enlarged view; (b) total thrust after binary discrimination and partially enlarged view; (c) total thrust after smoothing.

correlation between variables. The closer  $|\rho|$  is to 1, the stronger the correlation between variables, and vice versa. It can be seen from the correlation coefficient in the last column of Table 4 that the tunneling-specific energy, thrust factor, and average soil pressure are highly correlated with the strata. The cutter head torque and tunneling speed are strongly correlated with the strata.

Since the buried depth of the tunnel has a great influence on the average soil pressure, the average soil pressure is ignored. The correlation degree between the

tunneling parameters can be obtained by Pearson correlation analysis. The tunneling-specific energy is highly correlated with the thrust factor and has strong collinearity. It is difficult to accurately distinguish the influence of each variable on the formation recognition results by using both of them. Therefore, the tunneling-specific energy with a higher correlation with the formation is selected. Finally, the tunneling parameter vectors  $(T, v)$ ,  $(T, SE)$ ,  $(v, SE)$ , and  $(T, v, SE)$  composed of cutter head torque  $T$ , tunneling speed  $v$ , and tunneling specific energy

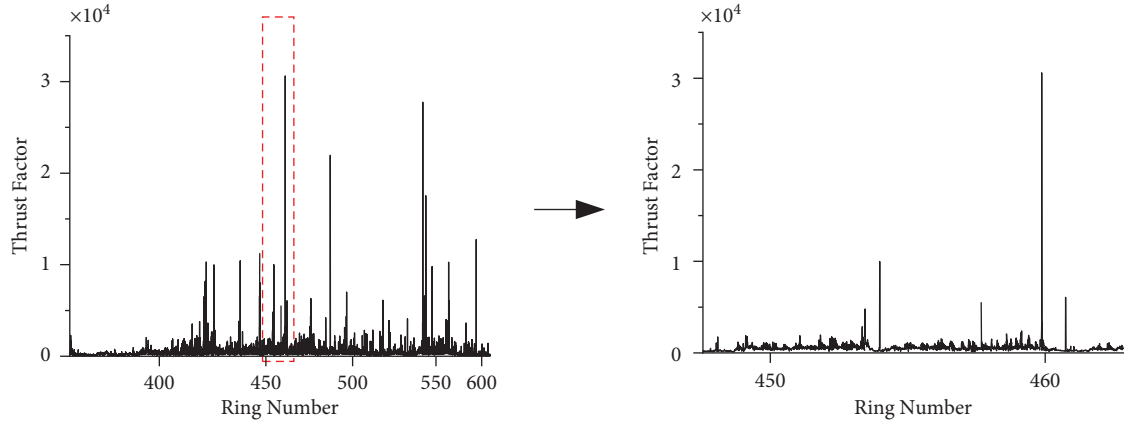


FIGURE 4: Time domain diagram of thrust factor and partially enlarged view.

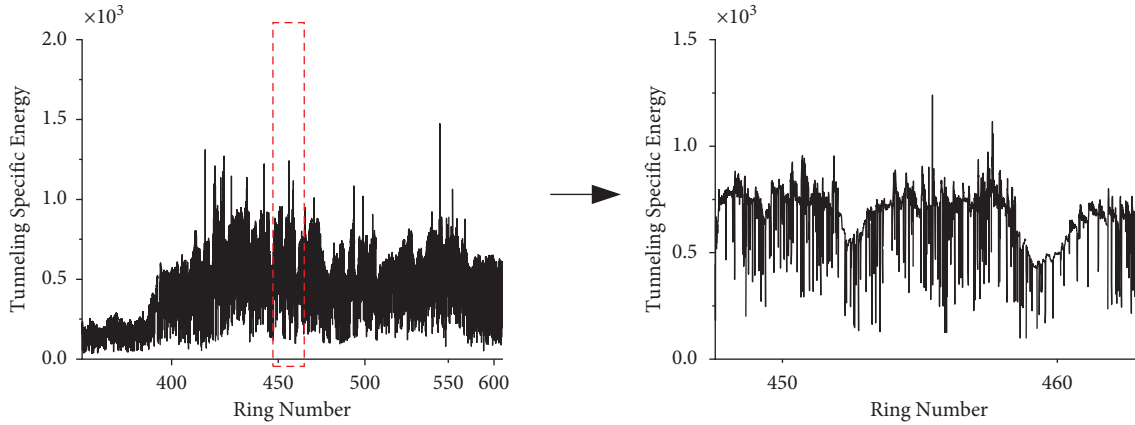


FIGURE 5: Time domain diagram of tunneling specific energy and partially enlarged view.

TABLE 4: Pearson correlation analysis.

Factor	Cutter head torque	Speed of tunneling	Average soil pressure	Thrust factor	Specific energy	Strata
Cutter head torque	1	0.438	0.610	0.755	0.730	0.725
Speed of tunneling		1	0.663	0.424	0.605	0.637
Average soil pressure			1	0.787	0.828	0.822
Thrust factor				1	0.955	0.911
Specific energy					1	0.947
Strata						1

$SE$  are formed, which are used as the data basis for stratum recognition based on K-nearest neighbor algorithm.

### 3. Formation Recognition Model Based on the K-Nearest Neighbor Algorithm

**3.1. K-Nearest Neighbor Algorithm.** As an online classification technology, K-Nearest Neighbor (KNN) algorithm is first applied to text classification research because of its simple theory, high accuracy, and good tolerance to outliers and noises. It is now widely used in classification and recognition fields such as face recognition and network public

opinion analysis [23, 24]. This paper applies the KNN algorithm model to stratum recognition.

Define  $U$  as training set data,  $n$  as a training set data quantity, then the training set is composed of cutter head torque  $U_1 = \{u_{11}, u_{12}, \dots, u_{1n}\}$ , tunneling speed  $U_2 = \{u_{21}, u_{22}, \dots, u_{2n}\}$ , and tunneling specific energy  $U_3 = \{u_{31}, u_{32}, \dots, u_{3n}\}$ .  $X$  is the test set data,  $m$  is the test set data quantity, then the test set is composed of cutter head torque  $X_1 = \{x_{11}, x_{12}, \dots, x_{1m}\}$ , tunneling speed  $X_2 = \{x_{21}, x_{22}, \dots, x_{2m}\}$ , and tunneling specific energy  $X_3 = \{x_{31}, x_{32}, \dots, x_{3m}\}$ . The distance set between training set data points  $U(u_{1i}, u_{2i}, u_{3i})$  and test set data points  $X(x_{1j}, x_{2j}, x_{3j})$  is  $D_{ij} = D(U(u_{1i}, u_{2i}, u_{3i}), X(x_{1j}, x_{2j}, x_{3j}))$ ,  $i \in (1, n)$ ,  $j \in (1, m)$ , and Euclidean distance is selected as the distance metric, as shown in the following equation:

$$D(U(u_{1i}, u_{2i}, u_{3i}), X(x_{1j}, x_{2j}, x_{3j})) = \sqrt{(u_{1i} - x_{1j})^2 + (u_{2i} - x_{2j})^2 + (u_{3i} - x_{3j})^2}, \quad (4)$$

$C_i$  is the classification attribute corresponding to  $U(u_{1i}, u_{2i}, u_{3i})$ , namely, strata. The neighborhood of the distance  $D$  between the element  $x$  of the test set  $X$  and the training set  $U$  is defined as  $\varepsilon$ , and the number  $n_\varepsilon$  of the neighborhood  $\varepsilon$  is defined in the following equation:

$$n_\varepsilon(x, U) = \{u_i \in U | D(x, u_i) \leq \varepsilon\}. \quad (5)$$

To make  $n_\varepsilon$  have at least  $k$ , the constraints shown in the following equation are added:

$$|n_\varepsilon(x, U)| \geq k. \quad (6)$$

If require that the number of neighborhoods  $\varepsilon'$  less than  $\varepsilon$  is not greater than  $k$ , add the constraints as shown in the following equation:

$$\forall \varepsilon' < \varepsilon, |n_{\varepsilon'}(x, U)| < k. \quad (7)$$

Two constraints of (6) and (7) make the number of distances  $n_\varepsilon$  nearest to  $x$  in  $U$  exactly  $k$ .

The set of  $k$  samples closest to  $x$  in  $U$  is defined as  $A_k$

$$A_k(x, U) = \{u_i \in U | D(x, u_i) \leq \varepsilon(k, x, U)\}. \quad (8)$$

Class  $c$  with the largest number of members in  $A_k$  is the final classification, and the classification result is represented by set  $P$ .

$$\% P(x, A_k(x, U)) = \arg \max \{ | \{u_i \in A_k(x, U) | C(u_i) = c \} | \}. \quad (9)$$

### 3.2. Selection of the $K$ Value Based on Cross-Test Method.

The training set data  $U$  is marked according to the category  $C$  of each data point, and the KNN algorithm is used to mark the category of each data point in the test set  $X$  in turn. The boundary between various categories in  $X$  forms the decision boundary. The decision boundary becomes smoother and smoother with the increase of the  $K$  value. If the  $K$  value is too small, the classification accuracy will be reduced. If  $K$  is too large and the samples in  $X$  are unbalanced, it will increase the noise and reduce the classification effect. Generally, the cross-test method is used to select the appropriate  $K$  value [25].

The test set data cannot be used to guide the training of the model. Therefore, the training set data are further divided into a training set and verification set (the training set data are divided into a training set and verification set according to 7:3 in this paper). Starting from selecting a small  $K$  value, the value of  $K$  is continuously increased, and the variance of the verification set is calculated. The verification set is used to evaluate the recognition rate of the KNN model under different  $K$  values. As shown in Figure 6, when  $K = 273$ , the recognition rate of the verification set  $R$  is

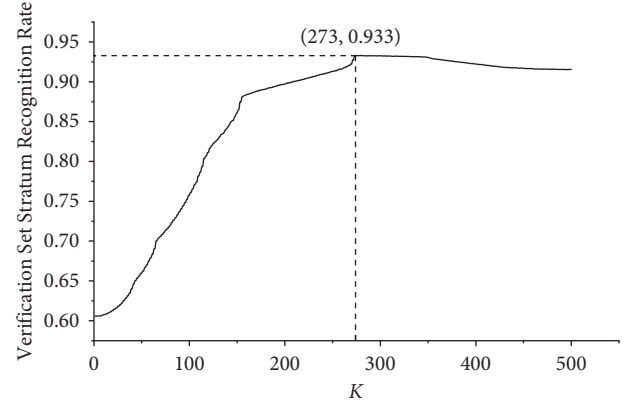


FIGURE 6: Relationship between the  $K$  value and formation recognition rate.

up to 93.3%, and the subsequent stratum recognition based on the KNN algorithm takes  $K = 273$ .

**3.3. Parameter Sensitivity Analysis.** Parameter sensitivity analysis of the formation recognition model is the change of recognition results caused by parameter changes, which is one of the important contents of model parameter uncertainty analysis, and also an indispensable part of the research and development and evaluation model [26–28]. Without losing generality, the probability density function of stratum identification is shown in the following equation:

$$q = f(x_1, x_2, \dots, x_i, x_n), \quad (10)$$

where  $q$  is the probability density of stratum recognition,  $x_i$  is the  $i$ th influencing factor,  $n$  is the number of factors affecting  $q$  ( $n = 3$  in this paper),  $x_1, x_2,$  and  $x_3$  are cutter head torque  $T$ , tunneling speed  $v$ , and tunneling specific energy  $SE$ , respectively.

When all the factors change from  $x_1, x_2,$  and  $x_3$  to  $x'_1, x'_2$  and  $x'_3$ , and their changes are  $\Delta x_1, \Delta x_2,$  and  $\Delta x_3$ , respectively,  $x'_1 = x_1 + \Delta x_1, x'_2 = x_2 + \Delta x_2, x'_3 = x_3 + \Delta x_3$ , the probability density function of stratigraphic recognition also changes from  $q$  to  $q'$ , and  $\Delta q = q' - q$  can be used to represent the change of  $q$  caused by all the factors. As shown in the following equation,  $\Delta q$  is expressed by Taylor expansion of multivariate function:

$$\Delta q = \frac{\partial q}{\partial x_1} \Delta x_1 + \frac{\partial q}{\partial x_2} \Delta x_2 + \frac{\partial q}{\partial x_3} \Delta x_3. \quad (11)$$

If only factor  $x_i$  changes, other factors do not change, that is,  $\Delta x_i \neq 0, \Delta x_l = 0, l \neq i$ , then the variation of the probability density  $q$  of stratigraphic recognition is denoted by  $\Delta q_i$ , as shown in the following formula:

$$\Delta q_i = \frac{\partial q}{\partial x_i} \Delta x_i, \quad i = 1, 2, 3 \quad (12)$$

The definition of sensitivity  $S_i$  is as follows:

$$S_i = \frac{\Delta q_i/q}{\Delta x_i/x_i} = \frac{[(\partial f/\partial x_i)/\Delta x_i]/q}{\Delta x_i/x_i} = \frac{\partial f}{\partial x_i} \times \frac{x_i}{q}. \quad (13)$$

The sensitivity of each tunneling parameter to the response result of the KNN algorithm stratum recognition model is shown in the following equation:

$$S_{xi} = \frac{\partial D}{\partial x_i} \cdot \frac{x_i}{D} = \frac{-x_i(u_i - x_i)}{D^2}, \quad i = 1, 2, 3. \quad (14)$$

$S_i > 0$  indicates that the change directions of  $\Delta q$  and  $\Delta x_i$  are the same. The larger  $|S_i|$  is, the more sensitive the probability density  $q$  of stratum identification is to the parameter  $x_i$ . The sensitivity of cutter head torque  $T$ , tunneling speed  $v$ , tunneling specific energy  $SE$ , and the sensitivity of multiple tunneling parameters ( $T, v$ ), ( $T, SE$ ), ( $v, SE$ ), and ( $T, v, SE$ ) is obtained, as shown in Figure 7.

It can be seen from Figure 7 that before the 450th ring, with the increase of data points, the sensitivity of each group of tunnel parameters to the stratum becomes higher and higher, and the sensitivity tends to be stable after the 450th ring. The sensitivity of  $v$ ,  $SE$  and  $[v, SE]$  is always low, which is not suitable for the input parameters of the subsequent KNN algorithm stratigraphic recognition model. The tunneling parameter vector  $[T, SE]$  has the highest sensitivity to stratum recognition. In the subsequent calculation,  $[T, SE]$  is used as the data basis for verifying and improving the KNN model.

**3.4. Verification of Stratum Recognition Model Based on the KNN Algorithm.** The cutter head torque  $T$  and tunneling specific energy  $SE$  constitute the tunneling parameter vector ( $T, SE$ ), which is used as the input of the KNN model for stratum recognition based on TBM tunneling parameters. Figure 8 shows the comparison between the strata identification results based on the KNN model and the actual strata.

It can be seen from Figure 8 that the identification error is mainly concentrated between the 277th ring and the 344th. After three changes in this interval, the identification error of the KNN model is relatively large. After the 351st ring, there are errors in the identification of the No. 5 formation, and the wrong data are concentrated in the same section. The main reason is that after calculating the  $K$  nearest distances in the KNN model, the stratum with the largest number of stratum types corresponding to these  $K$  distances is taken as the identification result. However, the data amount of tunneling parameters corresponding to stratum 1 and stratum 2 is relatively small, and the stratum changes many times in less data amount. It has certain limitations to rely solely on Euclidean distance as the basis for determining the stratum.

In 12936 test data, through the KNN model, 11765 data points are correctly identified, 1171 are wrongly identified, and the recognition rate is 90.95%. For 1171 false data, in order to improve the recognition rate of intervals with frequent formation changes, a formation recognition model of the  $K$ -nearest neighbor algorithm based on the entropy weight method is proposed.

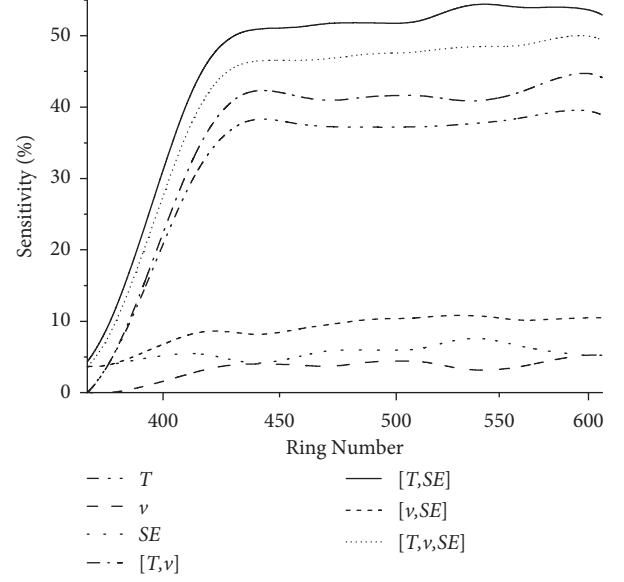


FIGURE 7: Sensitivity of tunneling parameters to strata identification model.

#### 4. Stratigraphic Recognition Model Based on K-Nearest Neighbor Algorithm of Entropy Weight Method

KNN algorithm calculates the  $K$  Euclidean distances closest to the test set data, but the density and importance of the type distribution of these  $K$  samples are different. Only taking the type with the largest number of  $K$  distances as the final judgment result will lead to large errors. Considering the influence of the density and importance of sample distribution on stratigraphic recognition, it is taken as the weight of distance. Under the premise of existing data, it is more reliable to carry out weight analysis from the information contained in the data itself. Therefore, the KNN algorithm based on the entropy weight method is proposed. Aiming at the problem of sample imbalance, the weight is determined by the difference in information content between samples and the accuracy of stratum recognition is improved.

**4.1. Stratigraphic Recognition Model Based on the K-Nearest Neighbor Algorithm of Normal Entropy Weight Method (EWM-KNN).** Entropy Weight Method (EWM), as a method to determine the weight through the amount of information, has strong objectivity and adaptability. It is mainly used to solve evaluation problems and avoid errors caused by artificial weighting [29]. This paper is mainly used for the weight analysis of the information contained in  $K$  points closest to the test set. In  $K$  points, the data distribution of each stratum type is weighted and scored by the entropy weight method, and the stratum type with the highest score is obtained [30, 31].

The information entropy of event  $X$  is shown in the following equation:



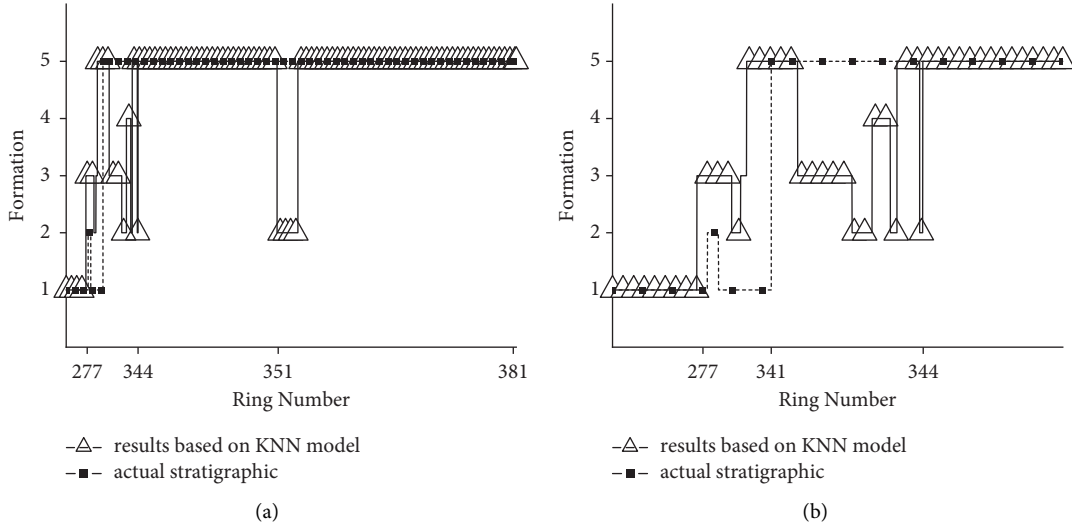


FIGURE 8: Formation recognition results based on the KNN model and actual stratigraphic correlation map. (a) Formation identification results of KNN model to test set. (b) Partially enlarged view.

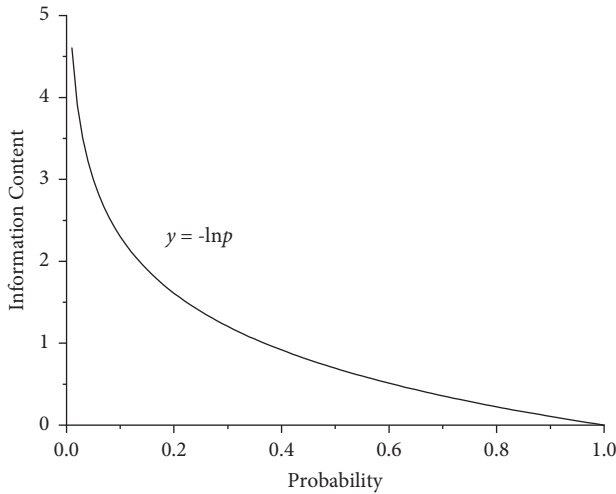


FIGURE 9: Diagram of the relationship between probability and information content.

$$H(x) = -\sum_{i=1}^n p(x_i) \ln p(x_i), \quad (15)$$

where  $x_i$  denotes the possible occurrence of event  $X$ ,  $p(x_i)$  is the probability of each occurrence, and  $-\ln p(x_i)$  denotes the amount of information contained in each case, their relationship is shown in Figure 9.  $-p(x_i) \ln p(x_i)$  is the expected value of information content. Therefore, the essence of information entropy is to represent the expected value of information. The maximum value of  $H(x)$  is  $\ln(n)$ .

For the  $i$ th distance information in  $K$  distances obtained by the KNN algorithm, divided by a constant  $\ln(n)$  the range is  $[0, 1]$ . The information entropy calculation is shown in the following equation:

$$e_j = \frac{1}{\ln(n)} \sum_{i=1}^n p_{ij} \ln p_{ij}, \quad j = 1, 2, 3, \dots, m. \quad (16)$$

The greater the value of information entropy, the smaller the amount of information. Therefore,  $w_j = 1 - e_{ij}$  is used as the utility value of information. After normalization, the entropy weight of distance information is obtained. The calculation formula is as follows:

$$W_j = \frac{w_j}{\sum_{j=1}^m w_j}, \quad j = 1, 2, 3, \dots, m. \quad (17)$$

As an online classification technology, EWM-KNN includes the advantages of the KNN algorithm, chooses tanh as the activation function [32], and improves the accuracy of formation identification through the entropy weight method (EWM).

**4.2. Verification of Stratum Recognition Model Based on Entropy Weight K-Nearest Neighbor Algorithm (EWM-KNN).** The  $[T, SE]$  tunneling parameter vector is used as input, and 1171 data recognition errors are found in 12936 test data based on the traditional KNN model. By reidentifying, the 1171 test data with the formation identification model based on the entropy weight K-nearest neighbor algorithm (EWM-KNN), the purpose of identifying the interval with more frequent formation changes is achieved, so as to improve the formation identification rate. Figure 10 shows the comparison between the stratum identification results based on the EWM-KNN model and the actual stratum.

It can be seen from Figure 10(b) that the identification error of the EWM-KNN model is mainly concentrated between the 277th ring and 341st ring, which realizes the identification of No. 2 formation with the least sample size, but there are still mainly concentrated in the identification of formation No. 1 to No. 5. The possible reasons for the identification error are that the sample size of the stratum core is small, and the fitting error of the stratum data and the tunneling parameters is large.

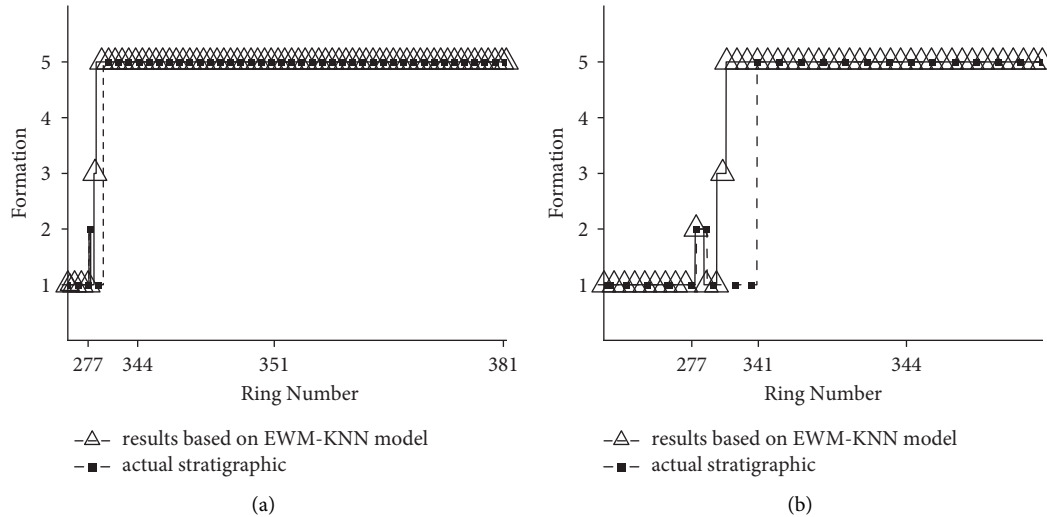


FIGURE 10: Formation recognition results based on the EWM-KNN model and actual stratigraphic correlation map. (a) Formation identification results of EWM-KNN model to test set. (b) Partially enlarged view.

In 1171 test data, through the EWM-KNN model, 984 data points are correctly identified, 187 data are wrongly identified, and the recognition rate is 84.03%. Combined with the previous recognition results, 12749 data points were correctly identified in 12936 test data, and the improved recognition rate was 98.55%. The entropy weight method is used to solve the problem of low recognition rate in the frequent interval of strata change, and the EWM-KNN model is used to improve the recognition rate of strata.

## 5. Discussion

The general KNN algorithm has been widely used in the classification and recognition fields such as face recognition and network public opinion analysis. However, as an inert learning algorithm, the KNN algorithm constructs the model at the last moment of the classification of the given test set and has certain requirements for the calculation time and equipment storage space (Section 3.1). Because the algorithm is based on Euclidean distance, but the Euclidean distance is not scale-invariant, that is, the calculated distance may be skewed according to the unit of the elements. Generally, data need to be normalized before using this distance measure. With the increase of data dimension, Euclidean distance loses its physical meaning, making the distance in a high dimension very unintuitive. However, in addition to the Euclidean distance, other commonly used distance measurement methods (Cosine similarity, Manhattan distance, Chebyshev distance, Jaccard index, and Haversine distance) have their own application fields and applicable conditions, which are not suitable for the research method in this paper. Therefore, when the KNN algorithm is adopted in this paper, the measurement content of Euclidean distance is improved [23, 24].

There is also an important parameter in the KNN algorithm, that is, the selection of the  $K$  value. How many

points should be extracted closest to the test set as the final classification set is crucial to the final recognition results? Therefore, the cross-validation method is used to select the  $K$  value (Section 3.2). The commonly used cross-validation forms include holdout verification,  $K$ -fold cross-validation, and leave-one-out cross-verification (LOOCV). Among them, Holdout Verification is not a strict cross-validation, because the data is not cross-used. The randomly selected part of the initial sample forms cross-validation data, and the remaining data as training data. This makes the test set used in advance, resulting in errors in subsequent recognition results. LOOCV uses one of the original samples as validation data, and the remaining variables are left as training data. This makes the factors affecting the results not fully considered, and it is also easy to lead to identification errors [25]. Therefore,  $K$ -fold cross-validation is selected, and 30% of the training set is taken as the validation set to determine the  $K$  value. The selection of the  $K$  value and the accuracy of model recognition are shown in Figure 6.

When the distance metric in the KNN algorithm is improved, the entropy weight method is used to weigh the Euclidean distance (Section 4.1). The data generated in the process of TBM tunneling is random. The method of subjective weighting requires the weight setter to have rich experience and rely too much on human factors [30, 31]. The objective weighting method determines the weight of parameters according to the correlation degree of each parameter attribute, or the amount of information provided by each parameter. It has strong objectivity and a theoretical basis. It is not only limited to the same project or TBM, but also can be widely used in the stratum identification of TBM tunnels, and has strong universality. The relationship between the probability of event occurrence and the amount of information is shown in Figure 9. However, the weight determined by the objective weighting method is likely to be inconsistent with the actual situation. Therefore, when

selecting the tunneling parameters involved in stratum recognition, the correlation analysis [13] (Section 2.3, Table 3) and sensitivity analysis [26–28] (Section 3.3, Figure 7) are used to eliminate the redundant parameters with overlapping information, reduce the dimension of input parameters of stratum recognition model, and avoid the interference between information.

## 6. Conclusions

This paper provides a new method of identifying stratum through machine learning using shield tunneling parameters. Its steps include preprocessing and feature extraction of tunneling data, digitization of geological features, and optimization of the stratum recognition model. The model is tested by the tunneling parameters of shield tunneling projects in Guangdong. The following conclusions are drawn by comparing the recognition effects of the two models:

- (1) The parameters  $T$  and SE that can best reflect the stratum in the composite stratum are extracted, which reduces the dimension of stratum recognition features and reduces the influence of redundant features on the accuracy of model recognition.
- (2) The KNN model determines the stratum to which the test data belongs by comparing the distance between the test data and the training data. The recognition rate of the whole stratum is 90.95%, and the recognition effect of the single stratum is good.
- (3) For the section with large soil viscosity and boulders, the recognition rate of the two models is about 8% higher than that of the KNN model alone. EWM-KNN model has a higher accuracy rate for stratum identification than the KNN model, and the identification accuracy rate can reach 98.55%, which is suitable for the identification of composite stratum.

## Data Availability

The (txt) data used to support the findings of this study have been deposited in the (4TU ResearchData) repository (DOI: 10.4121/19672173).

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Authors' Contributions

W. W., J. G., and J. L. conceptualized the study; W. W. developed methodology and formally analyzed the data; J. S., H. Q., and X. C. collected data resources; W. W. prepared writing of the original draft; W. W., J. G., and J. L. reviewed and edited the manuscript; J. G. supervised the study. All authors have read and agreed to the published version of the manuscript.

## Acknowledgments

This research was funded in part by the Graduate Innovative Research Fund of Hebei Province (grant no. CXZZBS202111) and National Key Research and Development Program (grant no. 2020YFB1709504).

## References

- [1] Q. Qian, "Main developments and directions of geological prediction and informatized technology of tunnel construction," *Tunnel Construction*, vol. 37, no. 3, pp. 251–263, 2017.
- [2] Q. Gong, L. Yin, H. Ma, and J. Zhao, "TBM tunnelling under adverse geological conditions: an overview," *Tunnelling and Underground Space Technology*, vol. 57, no. 3, pp. 4–17, 2016.
- [3] R. Hasanpour, J. Schmitt, Y. Ozcelik, and J. Rostami, "Examining the effect of adverse geological conditions on jamming of a single shielded TBM in Uluabat tunnel using numerical modeling," *Journal of Rock Mechanics and Geotechnical Engineering*, vol. 9, no. 6, pp. 1112–1122, 2017.
- [4] T. Yan, S. Shen, and A. Zhou, "Identification of geological characteristics from construction parameters during shield tunneling," *Acta Geotechnica*, vol. 18, pp. 1–17, 2022.
- [5] H. Chen and S. Liu, "Advanced geological prediction technology of tunnel based on image recognition," *Arabian Journal of Geosciences*, vol. 12, no. 19, p. 601, 2019.
- [6] K. Zhang, D. D. Tannant, W. Zheng, S. Chen, and X. Tan, "Prediction of karst for tunnelling using fuzzy assessment combined with geological investigations," *Tunnelling and Underground Space Technology*, vol. 80, pp. 64–77, 2018.
- [7] E. F. Salmi and E. J. Sellers, "A review of the methods to incorporate the geological and geotechnical characteristics of rock masses in blastability assessments for selective blast design," *Engineering Geology*, vol. 281, Article ID 105970, 2021.
- [8] B. Zhao, J. Zhou, Z. Tan, and T. Li, "Variation of shield boring parameters and correlation analysis in mixed ground," *China Civil Engineering Journal*, vol. 50, pp. 140–144, 2017.
- [9] E. Ghasemi, S. Yagiz, and M. Ataei, "Predicting penetration rate of hard rock tunnel boring machine using fuzzy logic," *Bulletin of Engineering Geology and the Environment*, vol. 73, no. 1, pp. 23–35, 2014.
- [10] A. Delisio and J. Zhao, "A new model for TBM performance prediction in blocky rock conditions," *Tunnelling and Underground Space Technology*, vol. 43, pp. 440–452, 2014.
- [11] D. Guo, J. Li, X. Li, Z. Li, P. Li, and Z. Chen, "Advance prediction of collapse for TBM tunneling using deep learning method," *Engineering Geology*, vol. 299, Article ID 106556, 2022.
- [12] K. Elbaz, T. Yan, A. Zhou, and S. L. Shen, "Deep learning analysis for energy consumption of shield tunneling machine drive system," *Tunnelling and Underground Space Technology*, vol. 123, Article ID 104405, 2022.
- [13] J. Liu, J. Guo, and X. Wang, "Recognition methods for boulder geology based on shield tunneling parameters," *Tunnel Construction*, vol. 39, no. 7, pp. 1132–1140, 2019.
- [14] X. Huang and C. She, "Research on methods of surrounding rock masses stability classification based on extension theory," *Rock and Soil Mechanics*, vol. 27, no. 10, pp. 1800–1804, 2006.
- [15] S. Hou, Y. Liu, and Q. Yang, "Real-time prediction of rock mass classification based on TBM operation big data and stacking technique of ensemble learning," *Journal of Rock Mechanics and Geotechnical Engineering*, vol. 14, no. 1, pp. 123–143, 2022.

- [16] Q. Liu, X. Wang, X. Huang, and X. Yin, "Prediction model of rock mass class using classification and regression tree integrated AdaBoost algorithm based on TBM driving data," *Tunnelling and Underground Space Technology*, vol. 106, Article ID 103595, 2020.
- [17] T. Yan, S. L. Shen, A. Zhou, and X. Chen, "Prediction of geological characteristics from shield operational parameters by integrating grid search and K-fold cross validation into stacking classification algorithm," *Journal of Rock Mechanics and Geotechnical Engineering*, vol. 14, no. 4, pp. 1292–1303, 2022.
- [18] H. Jin, W. Gao, X. Ma et al., *GB 50307-2012, Code for Geotechnical Investigations of Urban Rail Transit*, Standardization Administration of the P. R. C, Beijing, China, 2012.
- [19] Y. Teng, X. Huang, S. Wang et al., *GB 50007-2011, Code for Design of Building Foundation*, General Administration of Quality Supervision, Inspection and Quarantine of the P. R. C, Beijing, China, 2011.
- [20] Y. Zhang, Z. Cai, Y. Leng, and Y. Kang, "Correlative analysis of shield tunneling data and recognition of geologic features," *Journal of Harbin Engineering University*, vol. 32, no. 4, pp. 476–480, 2011.
- [21] J. Hassanpour, V. A. A. Ghaedi, J. Rostami, and A. Cheshomi, "Evaluation of common TBM performance prediction models based on field data from the second lot of Zagros water conveyance tunnel," *Tunnelling and Underground Space Technology*, pp. 147–156, 2016.
- [22] E. Chen, S. Deng, F. Chen, and C. Ma, "Development of a novel rock strength estimation model based on TBM boring performance," *Journal of Shandong University*, vol. 47, no. 2, pp. 7–13, 2017.
- [23] T. Abeywickrama, M. A. Cheema, and D. Taniar, "k-Nearest Neighbors on road networks: a Journey in experimentation and in-memory implementation," *Proceedings of the VLDB Endowment*, vol. 9, no. 6, pp. 492–503, 2016.
- [24] X. Yin, Q. Liu, X. Huang, and Y. Pan, "Perception model of surrounding rock geological conditions based on TBM operational big data and combined unsupervised-supervised learning," *Tunnelling and Underground Space Technology*, vol. 120, Article ID 104285, 2022.
- [25] E. Vasaneli, F. Micelli, D. Colangiuli, A. Calia, and M. A. Aiello, "A non destructive testing method for masonry by using UPV and cross validation procedure," *Materials and Structures*, vol. 53, no. 6, p. 134, 2020.
- [26] P. A. Vanrolleghem, G. Mannina, A. Cosenza, and M. B. Neumann, "Global sensitivity analysis for urban water quality modelling: terminology, convergence and comparison of different methods," *Journal of Hydrology*, vol. 522, pp. 339–352, 2015.
- [27] H. Shi, J. Chen, L. Hu, S. Geng, T. Zhang, and Y. Feng, "Multi-parameter sensitivity analysis on thermal characteristics of stratospheric airship," *Case Studies in Thermal Engineering*, vol. 25, Article ID 100902, 2021.
- [28] Z. Yuan, P. Liang, T. Silva, K. Yu, and J. E. Mottershead, "Parameter selection for model updating with global sensitivity analysis," *Mechanical Systems and Signal Processing*, vol. 115, pp. 483–496, 2019.
- [29] R. Kumar, S. Singh, P. S. Bilga et al., "Revealing the benefits of entropy weights method for multi-objective optimization in machining operations: a critical review," *Journal of Materials Research and Technology*, vol. 10, pp. 1471–1492, 2021.
- [30] Z. Wu and S. Zou, "A static risk assessment model for underwater shield tunnel construction," *Sādhanā*, vol. 45, no. 1, p. 215, 2020.
- [31] Y. Xue, H. Liu, C. Bai et al., "Extension prediction model of soft rock tunnel deformation grade based on entropy weight method and rough set," *Environmental Earth Sciences*, vol. 81, no. 1, p. 24, 2022.
- [32] S. L. Shen, N. Zhang, A. Zhou, and Z. Y. Yin, "Enhancement of neural networks with an alternative activation function tanhLU," *Expert Systems with Applications*, vol. 199, Article ID 117181, 2022.