

## Article

# Predicting Model of Dual-Mode Shield Tunneling Parameters in Complex Ground Using Recurrent Neural Networks and Multiple Optimization Algorithms

Taihua Yang<sup>1</sup>, Tian Wen<sup>1</sup>, Xing Huang<sup>2,\*</sup>, Bin Liu<sup>2,\*</sup>, Hongbing Shi<sup>3</sup>, Shaoran Liu<sup>4</sup>, Xiaoxiang Peng<sup>1</sup> and Guangzu Sheng<sup>5</sup>

<sup>1</sup> School of Urban Construction, Wuhan University of Science and Technology, Wuhan 430083, China; yangtaihua@wust.edu.cn (T.Y.)

<sup>2</sup> State Key Laboratory of Geomechanics and Geotechnical Engineering, Institute of Rock and Soil Mechanics, Chinese Academy of Sciences, Wuhan 430071, China

<sup>3</sup> China Construction Civil Infrastructure Corp., Ltd., Beijing 100029, China

<sup>4</sup> China Construction South Investment Co., Ltd., Shenzhen 518000, China

<sup>5</sup> Wuhan Urban Construction Group Construction Management Co., Ltd., Wuhan 430040, China

\* Correspondence: xhuang@whrsm.ac.cn (X.H.); liubin@whrsm.ac.cn (B.L.)

**Abstract:** Based on the left tunnel of the Liuxiandong Station to Baimang Station section of Shenzhen Metro Line 13 (China), a prediction model for the advanced rate of dual-mode shield tunneling in complex strata was established to explore intelligent tunneling technology in complex ground. Firstly, geological parameters of the complex strata and on-site monitoring parameters of EPB/TBM dual-mode shield tunneling were collected, with tunneling parameters, shield tunneling mode, and strata parameters selected as input features. Subsequently, the Isolation Forest algorithm was employed to remove outliers from the original advance parameters, and an improved mean filtering algorithm was applied to eliminate data noise, resulting in the steady-state phase parameters of the shield tunneling process. The base model was chosen as the Long-Short Term Memory (LSTM) recurrent neural network. During the model training process, particle swarm optimization (PSO), genetic algorithm (GA), differential evolution (DE), and Bayesian optimization (BO) algorithms were, respectively, combined to optimize the model's hyperparameters. Via rank analysis based on evaluation metrics, the BO-LSTM model was found to have the shortest runtime and highest accuracy. Finally, the dropout algorithm and five-fold time series cross-validation were incorporated into the BO-LSTM model, creating a multi-algorithm-optimized recurrent neural network model for predicting tunneling speed. The results indicate that (1) the Isolation Forest algorithm can conveniently identify outliers while considering the relationship between tunneling speed and other parameters; (2) the improved mean filtering algorithm exhibits better denoising effects on cutterhead speed and tunneling speed; and (3) the multi-algorithm optimized LSTM model exhibits high prediction accuracy and operational efficiency under various geological parameters and different excavation modes. The minimum Mean Absolute Percentage Error (MAPE) prediction result is 8.3%, with an average MAPE prediction result below 15%.

**Keywords:** shield tunneling; complex strata; EPB/TBM dual-mode shield tunneling; tunneling parameter prediction; recurrent neural network



**Citation:** Yang, T.; Wen, T.; Huang, X.; Liu, B.; Shi, H.; Liu, S.; Peng, X.; Sheng, G. Predicting Model of Dual-Mode Shield Tunneling Parameters in Complex Ground Using Recurrent Neural Networks and Multiple Optimization Algorithms. *Appl. Sci.* **2024**, *14*, 581. <https://doi.org/10.3390/app14020581>

Academic Editor: José António Correia

Received: 7 December 2023

Revised: 26 December 2023

Accepted: 29 December 2023

Published: 9 January 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The composite stratigraphy dual-mode shield tunnel boring machine (TBM) is a shield construction technology designed to cope with complex geological conditions in subway tunneling. It can efficiently and stably advance in composite strata, such as alternating soft and hard layers or soft overlying complex layers. Among these, the most common is the EPB/TBM dual-mode shield TBM. Compared to traditional single-mode shield TBMs, the

composite stratigraphy dual-mode shield TBM can better adapt to geological conditions, enhancing construction efficiency and safety. When using a dual-mode shield TBM for variable excavation modes in complex geological conditions [1–3], it is necessary to comprehensively consider factors such as engineering geology and excavation mode and propose reasonable adjustments to the equipment excavation parameters. Optimizing the shield TBM excavation parameters is crucial for ensuring stable and effective tunneling [4]. Improper excavation parameter settings may lead to various issues, such as over-excavation, deviation of the TBM axis, instability of the face support [5], excessive tool wear, and in severe cases, may even result in TBM accidents. Therefore, in the shield construction process, careful attention must be paid to the setting and adjustment of excavation parameters, and strict adherence to relevant construction regulations and standards is essential to ensure the safety and efficiency of shield construction [6–8]. During the construction of composite stratigraphy shield tunnels, the tunneling speed of the TBM may be influenced by various factors, including geological conditions, shield construction parameters, and environmental factors [9–11]. The relationship between these factors and the TBM's tunneling speed involves complex mathematical relationships, requiring further in-depth research and analysis. Precise control and adjustment of these factors are necessary to ensure smooth construction.

Early tunneling speed prediction models are primarily based on empirical, theoretical, or a combination of both approaches. These models explore the operational patterns of tunneling speed and other characteristics via formula derivation and simulated experiments. Zhao Bojian et al. [12] utilized statistical methods to establish relationships between shield tunneling parameters and strata, conducting a thorough analysis. In the context of composite strata, Li Jie et al. [13] employed orthogonal experiments combined with nonlinear regression analysis to develop a mathematical model for the tunneling speed of Earth Pressure Balance (EPB) shield tunneling. Sapigni et al. [14], via a study of monitoring data from three tunnels, found a close correlation between excavation rate and rock mass classification, which could be fitted with a quadratic regression equation. Based on on-site measured data, Kahrama et al. [15] established a regression model for excavation rate, with statistical analysis results indicating a close correlation with rock properties. Hasanpour et al. [16], using data from the Nowsood Tunnel No. 2, established an empirical formula for excavation rate about different geological parameters, finding a particularly close correlation with rock cuttability, especially the field penetration index. Wang Hongxin et al. [17], based on model test results, successfully developed a structural model to study EPB shield tunneling. They derived specific mathematical expressions for the total thrust, soil chamber pressure, screw conveyor speed, and tunneling speed. Zhang Zhiqi et al. [18] conducted multivariate regression analysis and discovered a certain robustness between shield tunneling speed and cutterhead torque. Xu Qianwei [19], via experiments, identified key shield construction parameters and studied their adaptive relationships with soil properties. However, these methods rely on linear relationships between data, while shield tunneling data often exhibit nonlinear relationships. Additionally, shield tunneling data are exceptionally voluminous, making conventional calculation methods unable to calculate the connections between data precisely.

In recent years, due to the advancement of artificial intelligence technology, many data-driven models have been successfully applied to tunneling speed prediction. Xu et al. [20], based on on-site and laboratory experiment data from a tunnel in Malaysia, proposed five different machine-learning methods for predicting tunneling speed. By comparing the predicted results of each method with actual values, they found that the K-Nearest Neighbors (K-NN) algorithm achieved the best prediction accuracy. Zhang Zheming et al. [21] employed the uniform extraction of samples to establish a model for predicting cutterhead torque, cutterhead thrust, and tunneling speed in the stable section. The model used a radial basis function neural network kernel and ten-fold cross-validation for the training of the Least Squares Support Vector Machine (LS-SVM) model, providing accurate predictions. Based on the Shenzhen Metro project, Li Chao et al. [22] utilized

backpropagation (BP) artificial neural network technology to establish a prediction model for shield tunneling parameters under complex geological conditions. Hou Shaokang et al. [23] introduced a novel TBM tunneling parameter prediction model that used an improved particle swarm optimization algorithm to optimize a BP neural network. The enhanced algorithm employed adaptive inertia weights, resulting in higher prediction accuracy than traditional BP and PSO-BP models. Qiu Daohong et al. [24], considering the temporal nature of collected TBM tunneling parameters, constructed a Long Short-Term Memory (LSTM) neural network model. Experimental results demonstrated that this model achieved optimal prediction accuracy for net tunneling speed.

There is a greater abundance of research on the tunneling speed prediction of single-mode shield TBMs under different geological conditions. At the same time, there is relatively less research on the tunneling speed prediction of dual-mode shield TBMs. Dual-mode shield TBMs require the selection of different excavation modes based on varying geological conditions during the tunneling process. Simultaneously, factors such as face support balance, cutterhead tool types, and support methods may change, rendering the parameters obtained by the data acquisition system more complex. Therefore, establishing tunneling speed prediction models for dual-mode shield TBMs is inherently more complex than for single-mode shield TBMs. Furthermore, the application of machine learning algorithm models for predicting tunneling speed has been more prevalent in research. The above model data input, lack of data refinement process, data outlier processing, and its importance. In contrast, there is a relatively limited number of studies focusing on prediction models based on optimization algorithms. There is also a scarcity of research that compares the impact of different optimization algorithms on the prediction accuracy of models.

Based on the left line project of Liuxiandong Station–Baimang Station Tunnel of Shenzhen Metro Line 13, this paper obtains a large number of time series characteristic parameters based on the data acquisition system, eliminates abnormal data via the isolated forest algorithm, and optimizes the original shield parameters with the improved mean filtering algorithm. Considering the influence of stratum conditions on tunneling parameters from three dimensions of surrounding rock grade, tunnel depth–span ratio, and soft–hard composite ratio, an LSTM model integrating four super-parameter optimization algorithms is established. Combined with the dropout algorithm and five-fold time series cross-validation, the two shield tunneling modes of EPB and TBM and the propulsion speed under different strata are predicted and analyzed, which provides feasible guidance for intelligent control of the dual-mode shield tunneling process.

## 2. Project Overview

The Shenzhen Metro Line 13 is a north–south urban rail transit line starting at the Shenzhen Bay Port and traversing the Nanshan District and Bao’an District, totaling 22,434 km. The line is between the central development axis and the western development axis of the city, connecting the Hohai Central Urban Area and the Western High-Tech Industrial Park. It serves as a fast connection between these two regions. This study is based on the left-line project from Liuxiandong Station to Baimang Station, focusing on the section from 0 to 650 m in the tunnel, with a total length of approximately 2036 m. This section represents a typical complex geological formation. The Liuxiandong Station’s elevation to Baimang Station’s section ranges from 24.26 to 42.85 m above ground, with a slightly undulating terrain. A mid-tableland landscape, with localized gully landscapes between tablelands, characterizes the initial topography. The predominant geological layers in this section include the fourth series of artificial fill (Q4ml) and fluvial deposits (Q4al + pl). The fluvial deposits mainly consist of silty clay, plastic fine-grained clay, and sand layers. The existing geological layers exhibit distinct stratified structures above the fill layer. The underlying bedrock consists mainly of mixed granites from the Jixian Formation to the Qingbaikou Formation (Jx-Qby) and biotite granite from the Yanshan period ( $\gamma\beta 5K1$ ). The tunnel body passes through moderately to slightly weathered rocks as the primary

geological formation, followed by residual layers and highly weathered rocks. The tunnel also intersects sporadically with intensely weathered and moderately weathered schists.

The shield tunneling machine selected for the Liubai section of Shenzhen Metro Line 13 is the “Xinhe 15” EPB and single-shield TBM dual-mode shield tunneling machine, designed and manufactured by China Railway Construction Equipment Group. The main configuration parameters are detailed in Table 1. The design concept of this shield tunneling machine integrates the functionalities of both EPB and single-shield TBM, allowing for in-tunnel conversion. The EPB mode is suitable for weak geological conditions, employing the EPB tunneling method to ensure the face support’s stability and prevent uneven ground settlement. On the other hand, the single-shield TBM mode applies to complex rock formations, enhancing the tunneling speed in such conditions and avoiding risks associated with slow progress and severe tool wear when using the EPB mode in complex rock formations.

**Table 1.** “Xinhe 15” dual-mode shield main technical parameters index.

Parameter	Design Value
Shield type	Mono-protecting shield
Total weight/T (Host + Supporting)	650
Total length of machine/m	96
Cutterhead speed/(r/min)	0-2.47-5.59
Excavation diameter/(mm)	φ6980
Maximum thrust/T	5060
Rated torque/kN·m	6080
Number of tools	50 (Hob) + 61 (Scraper) + 12 (Side scraper)

### 3. Predictive Model Algorithm Principle

#### 3.1. LSTM Model

The data of shield tunneling are time-series, and the data are purely dependent. Traditional machine learning algorithms (such as BP neural network, random forest, etc.) cannot capture the time-series value between data. The LSTM network has an internal gating mechanism, which enables it to effectively capture and retain information from past inputs. At the same time, the generalization ability of the LSTM neural network is stronger. It is a deep learning algorithm widely used in the market and more suitable for engineering needs. In predicting TBM tunneling speed, the LSTM model can be employed by inputting parameters such as excavation parameters, TBM mode, and geological parameters. This allows for the establishment of a model capable of predicting the tunneling speed at the next step. Furthermore, by introducing hyperparameter optimization algorithms into the LSTM model, prediction accuracy and operational efficiency can be enhanced, providing better support for TBM excavation projects. When handling sequential data, the LSTM model effectively addresses issues such as gradient vanishing or exploding that exist in traditional Recurrent Neural Networks (RNNs). Additionally, it can model long-term dependencies. In the LSTM architecture, besides the conventional input layer, output layer, and hidden layer, a memory cell and three gating units (Forget Gate, Input Gate, Output Gate) are introduced. Among these,  $i_t$  represents the input gate,  $f_t$  is the forget gate,  $g_t$  denotes the input supply, and  $o_t$  represents the output gate. The formulas for computing the forget gate, input gate, and output gate are as follows:

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

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i), \tag{2}$$

$$C_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C), \tag{3}$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o), \tag{4}$$

In the formula,  $W_f$ ,  $W_i$ ,  $W_c$ , and  $W_o$  are different calculation matrices;  $b_f$ ,  $b_i$ ,  $b_o$ , and  $b_c$  are the bias terms of the three gated units and the cell state, respectively;  $\sigma$  and  $\tanh$  are activation functions, respectively;  $\sigma$  represents the *sigmoid* function, and its output is between 0 and 1; and  $\tanh$  is the hyperbolic tangent function of the mapping interval  $[-1, 1]$ . At each time step  $t$ , the LSTM introduces a hidden state  $C$  (cell state) and employs three gates to control the content of the cell state. The first gate, the Forget Gate, determines how much information from the previous time step's cell state  $C_{t-1}$  needs to be retained in the current time step's cell state  $C_t$ . The second gate, the Input Gate, regulates how much information from the current time step's input  $x$  is stored in the cell state  $C_t$ . The Output Gate, the third gate, determines how much information will be output from the cell state  $C_t$  to the current time step's output  $h_t$ . The LSTM network can selectively retain or output helpful information via this mechanism, enabling improved handling of long sequential data.

### 3.2. Multiple Optimization Algorithm Model

In the operation of LSTM, performing hyperparameter optimization is crucial as it identifies the optimal combination of hyperparameters that enhances model performance. By selecting appropriate hyperparameter combinations, it is possible to improve the model's generalization ability, resulting in superior performance on the test set. This study employs four widely applicable and practical hyperparameter optimization algorithms: Genetic Algorithm (GA), Differential Evolution (DE), Bayesian Optimization (BO), and Particle Swarm Optimization (PSO). Due to constraints in length, the principles of these algorithms are not extensively elaborated. Subsequently, the study will train these four hyperparameter optimization algorithms, compare the performance of evaluation metrics under different algorithms, and employ a ranking method to determine the optimal hyperparameter optimization algorithm for this model.

After selecting the optimal hyperparameter optimization algorithm, in order to enhance the model's generalization ability, a multi-algorithm optimized model incorporating Time Series cross-validation (TSCV) and the Dropout algorithm was established to further improve the predictive performance of the model on tunneling speed. The Dropout algorithm is a widely used regularization technique for deep learning models proposed by Geoffrey Hinton and his team in 2012. Its primary objective is to prevent neural networks from overfitting, thereby improving the model's generalization ability across different data sets. This study incorporated the Dropout layer into the LSTM model, with a dropout probability set to 0.1. This implementation achieved random dropout of a portion of neurons in the neural network, preventing certain neurons from developing excessive dependence on specific features and thereby reducing the complexity of the neural network.

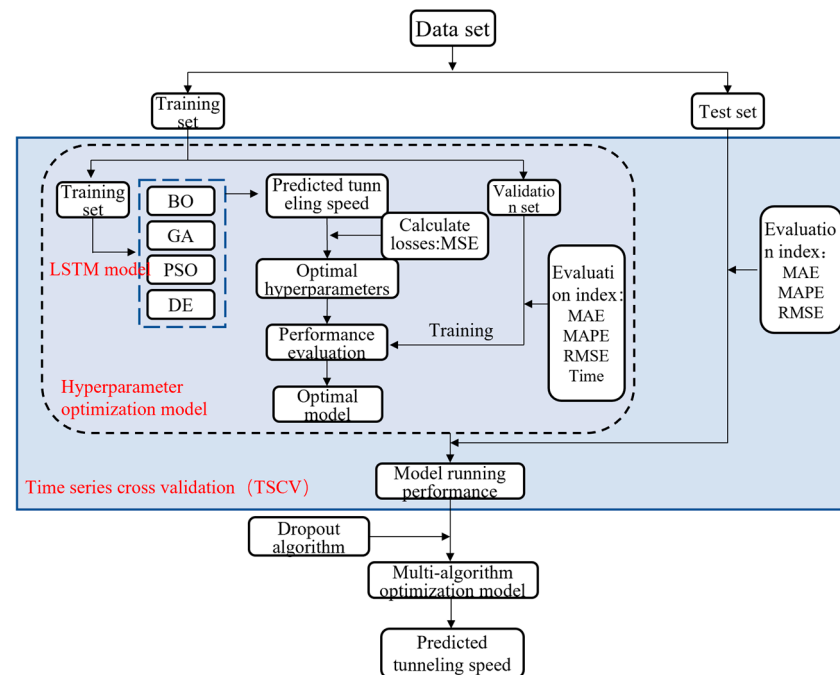
Tunnel boring machine (TBM) excavation parameters constitute time-series data with temporal dependencies. In conventional cross-validation, data is randomly partitioned, making validating past data using future data highly inappropriate due to the temporal dependencies. Therefore, the data sequence should not be arbitrarily shuffled. This study employs time series cross-validation to evaluate prediction models' performance on time-series data. When assessing time-series prediction models, time-series cross-validation effectively measures the model's generalization ability across different periods. The model utilizes a five-fold time series cross-validation, and the operational workflow is depicted in Figure 1:





**Figure 1.** Five-fold time series cross-validation diagram.

The process begins by selecting a window size, which encompasses a specific number of observational values. Subsequently, the window is incrementally moved forward, providing distinct training and validation data subsets for the model. At each step, the model is trained on the data within the window and validated on the data outside the window. This allows for the computation of performance metrics, such as mean squared and absolute percentage errors, for the model on each validation set. Finally, the average of these metrics is calculated to assess the model’s overall performance. By combining the above methods, the overall model process is shown in Figure 2.



**Figure 2.** Flow chart of multi-algorithm optimization model.

#### 4. Establishment of a Prediction Model for Dual-Mode Shield Tunneling Parameters

##### 4.1. Filter Input Feature Parameters

The predictive model in this study comprehensively considers the influence of various factors, including geological parameters, shield machine excavation parameters, and shield tunneling modes. Therefore, detailed explanations for the parameters mentioned above are provided in the following sections.

##### (1) Composite Geological Characteristics Parameters

When considering the geological conditions within the Liubai section’s left tunnel range, this study employs rock mass rating and the soft–hard compound ratio as in-

put parameters for composite geological features. According to the national standard 'Code for geotechnical engineering investigation of urban rail transit' (GB50307-2012) Appendix F [25], the geotechnical construction engineering classification of each rock and soil layer revealed by this investigation is carried out. Grade III rock mass predominates along the left tunnel route, accounting for 46.78% of the tunnel length. Grade IV rock mass represents 12.7% of the tunnel length, Grade V rock mass accounts for 38.56%, and Grade VI rock mass covers 1.95%. To better reflect the conditions of alternating soft and hard layers in the composite geological formation, the thickness of the weak soil layer at the excavation face to the total excavation face thickness is defined as the soft–hard compound ratio, representing the composite geological formation. The definition of the soft–hard compound ratio is given by Formula (5):

$$\mu = \frac{H_i}{H_i + H_j'} \quad (5)$$

The equation  $\mu$  represents the soft–hard compound ratio, within the range of 0 to 1;  $H_i$  is the thickness of the weak soil layer, and  $H_j'$  is the thickness of the hard rock layer.

#### (2) Operating parameters of shield machine

The original data in the research phase comprises 220-dimensional parameter indicators. Based on the mechanisms affecting shield tunneling speed, this study primarily considers several excavation parameters as input variables for predicting tunneling speed:

- (1) Total Thrust: A more significant total thrust reduces resistance encountered by the shield machine during excavation, enabling faster advancement. As the shield machine progresses, the total thrust must overcome the resistance and friction in the geological layers to propel the machine forward. Therefore, the magnitude of the total thrust directly influences the shield tunneling speed.
  - (2) Thrust Pressure: It represents the force exerted by the shield machine during the excavation process, directly affecting the machine's forward speed in the geological layers. The thrust pressure of the shield machine should be controlled within a specific range to ensure the stability and safety of shield tunneling.
  - (3) Cutterhead Torque: It represents the force exerted by the shield machine during the excavation process, directly affecting the machine's forward speed in the geological layers. The thrust pressure of the shield machine should be controlled within a specific range to ensure the stability and safety of shield tunneling.
  - (4) Cutterhead Speed: The higher the rotation speed of the cutterhead on the shield machine, the stronger its cutting ability, resulting in faster advancement of the shield machine in the geological layers.
- #### (3) Shield tunneling mode parameters

To consider variations in the tunneling state of the dual-mode shield machine, it is necessary to select different excavation modes based on distinct geological conditions and excavation performance. They need to be labeled accordingly to differentiate between the EPB and TBM excavation modes. After multiple test runs, the EPB mode is ultimately labeled 1, and the TBM mode is labeled 3. This labeling approach aids the model in learning from the input parameters, reduces the impact on neural network weight training, and consequently enhances the rationality of the model's predictive results.

### 4.2. Data Preprocessing

#### 4.2.1. Steady-State Segment Data Extraction

During the tunneling process, equipment monitors data on a timely basis. Interruptions in tunneling due to cutterhead tool replacement, segment assembly, and other reasons are within the scope of monitoring. This results in a substantial volume of raw data of lower quality. To extract steady-state data, it is necessary to eliminate empty thrusting data and exclude short-term unstable data [26]. In general processing, cutterhead thrust

(F), cutterhead torque (T), cutterhead speed (RPM), and tunneling speed (V) are considered state-discriminant parameters. Any value of zero parameters is considered blank data recorded during shield machine operations, and the entire row of data is removed. Short-term unstable data in each tunneling cycle, typically occurring during the start-up process of the shield machine, should be selectively excluded to reduce potential errors in subsequent calculations.

Figure 3 shows a shield tunneling process for data segment 140. Although the displayed unstable data in the initial stages of tunneling are minimal, considering that not every start-up process proceeds smoothly, it is advisable to sequentially exclude the first 10% of data from each start-up phase chronologically. This approach aims to optimize the quality of the data.

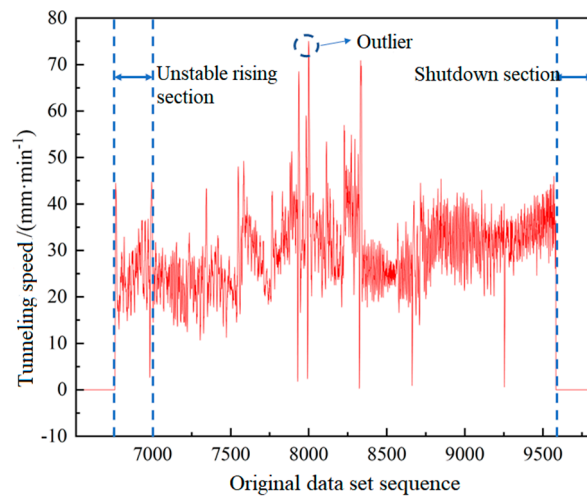


Figure 3. Raw data.

#### 4.2.2. Outlier Handling

In practical scenarios, equipment monitoring generates a large amount of data, and monitoring anomalies may inevitably lead to some outlier values. For such data, commonly used outlier detection methods are employed for exclusion. This study utilizes the Isolation Forest algorithm for identifying and detecting anomalous data. The Isolation Forest algorithm was jointly proposed by Professor Zhou Zhihua and others in 2008 [27] for data mining. It is an unsupervised anomaly detection algorithm suitable for continuous data anomaly detection. Specifically, the Isolation Forest algorithm randomly partitions the data into several subspaces, constructs a set of binary search trees based on random partitions, and inserts data points into the leaf nodes. It determines whether a data point is an outlier by calculating the average path length across all trees for each data point. A shorter path length indicates that the data point is more easily isolated, making it more likely to be an outlier. The isolated forest model is shown in Figure 4.

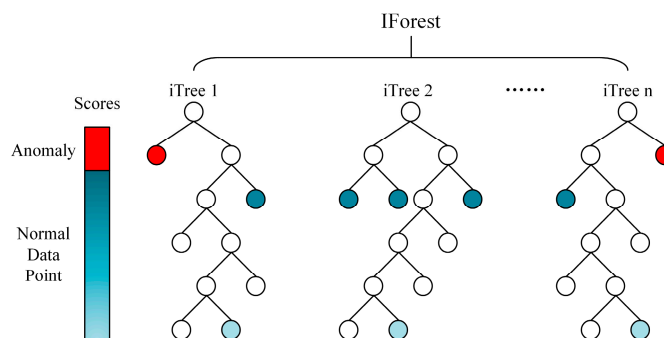


Figure 4. Isolated forest model.



The specific numerical expression is given by the following Formula (6). Firstly, construct  $h(x)$  as a metric to measure the “isolation degree” of a data point (sample) from other data points. It is defined as the path length that a data point traverses from the root node to that point on a random tree. For a random tree  $T$  and a data point  $x$  within it, the number of samples that share the same leaf node with  $x$  in  $T$  is denoted as  $T.size(x)$ . Subsequently, for a given  $T.size(x)$ , a correction term  $c(T.size(x))$  can be calculated to represent the average path length of constructing a binary tree with  $T.size(x)$  samples.

$$h(x) = e + c(T.size), \quad (6)$$

The second step involves constructing the average path length  $c(n)$  of a binary tree using  $n$  samples, specifically based on a Binary Search Tree (BST). This metric represents the average distance between any two nodes in a BST with  $n$  elements. The average path length  $c(n)$  can be calculated using the following formula:

$$c(n) = 2[H(n) - 1] - 2 \frac{n - q(n) - 1}{n}, \quad (7)$$

Here,  $H(n)$  represents the average path length of a BST with a height of  $n$ , and  $q(n)$  is an estimation of the number of non-leaf nodes in the BST after randomly constructing and inserting  $n$  elements.

$$s(x, n) = 2^{-\frac{E(h(x))}{c(n)}}, \quad (8)$$

Finally, a normalization process is applied to map the range of  $h(x)$  to between 0 and 1. Here,  $h(x)$  represents the path length of sample point  $x$ , and  $S(x, n)$  is the anomaly index of the tree, which records the training data of  $x$  in  $n$  samples. From this formula, it can be observed that as the path length decreases,  $s$  approaches 1, and the probability of detecting the sample point as an anomaly increases.

Compared to commonly used calculation methods such as Mahalanobis distance and the  $3\sigma$  criterion, the Isolation Forest algorithm does not require calculating anomaly standards for data under different geological conditions. Its unsupervised, efficient, and precise advantages make it more suitable for extensive data processing in industries such as tunnel boring machines (TBM). The Isolation Forest algorithm is capable of handling high-dimensional data. However, an increase in data dimensions during processing increases tree depth, resulting in higher time complexity for tree construction and search. Additionally, greater tree depth increases sensitivity to anomalies, making them more prone to be classified as outliers. Considering the focus of this study on predicting the tedious tunnel process's advance rate, the advance rate is paired with the total thrust, thrust pressure, cutterhead torque, and cutterhead speed to form a two-dimensional array. This array is then cyclically fed into the model for training, with all identified anomalies marked after training completion. Finally, the rows containing anomalies are removed. The model is configured with 100 trees, a contamination rate of 0.02 in the data set, and a random seed set to 42. The training results are shown in Figure 5.

For the processing of outliers in the 140th tunneling cycle parameters, a total of 2825 data sets were processed. During the operation, numerous duplicate data points were identified. Via statistical analysis, it was determined that a total of 115 data sets were flagged as outliers. As a result, the final data set comprises 2710 remaining data sets.

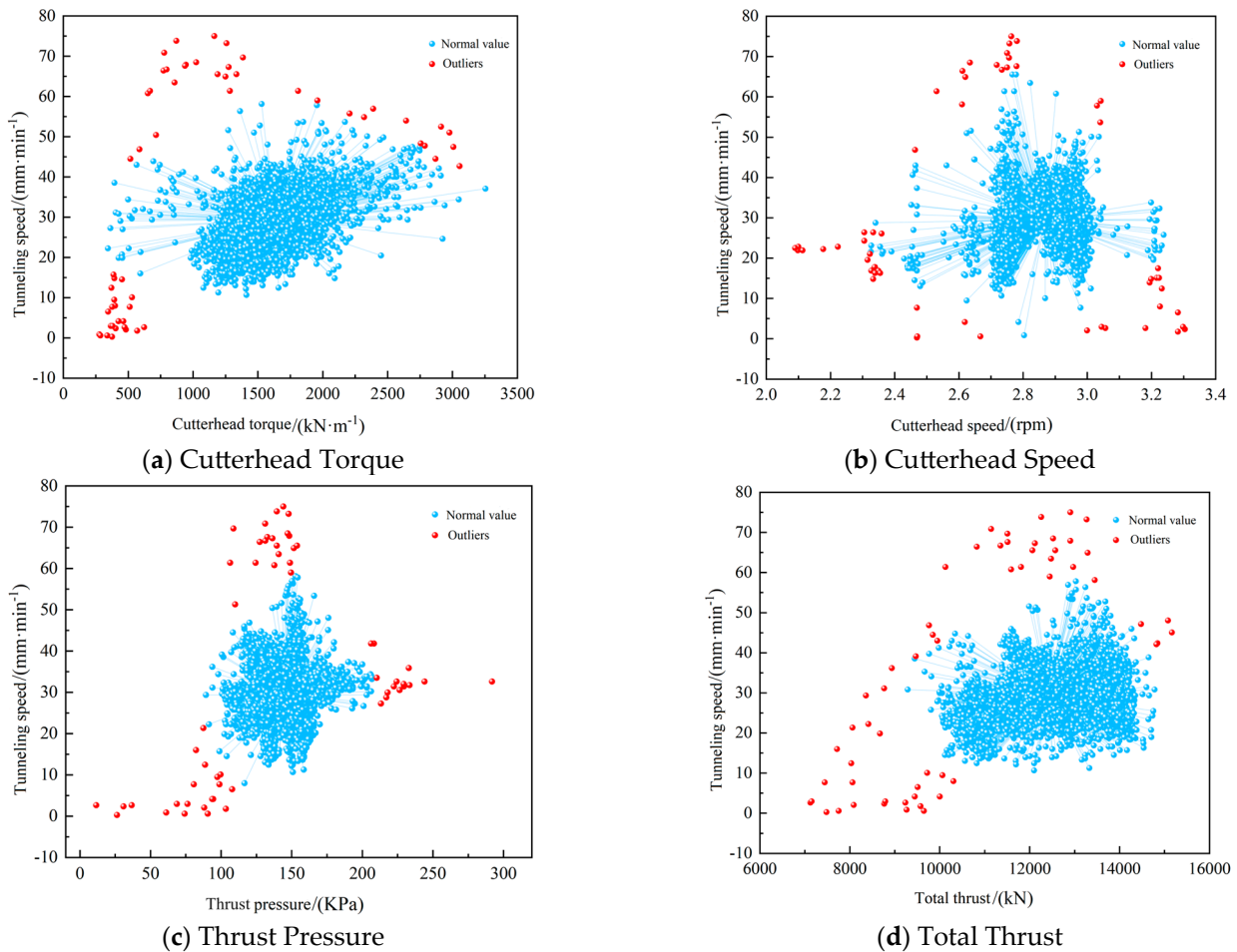


Figure 5. Outlier tag.

#### 4.2.3. Data Denoising

After removing outliers, the data exhibits inevitable fluctuations that can impact the modeling process and even the calculation results. These anomalous fluctuation data in the tunneling parameter sequence are called noise data. This phenomenon is particularly pronounced in geological environments with complex and variable conditions, rendering the temporal parameter data more unstable. Xiao et al. [28] pointed out that denoising tunneling parameters can reduce the spatial variability of these parameters, making it easier for machine learning algorithms to learn the patterns in data changes. An improved mean filtering algorithm is employed to denoise the tunneling parameters to mitigate these irregular variations. The formula for the original mean filtering algorithm is as follows, assuming a non-stationary data set of total length  $N$ . To eliminate noise, a window size of  $2n + 1$  ( $<N$ ) is set, and as the window slides forward, the average of every  $2n + 1$  adjacent data  $\bar{y}_k$  is taken to represent the measurement result of the midpoint data. This method effectively eliminates noise.

$$\bar{y}_k = \frac{1}{2n + 1} \sum_{i=-n}^{2n+1} y_{k+i} \tag{9}$$

There is significant data fluctuation with sudden rises and falls throughout the shield tunneling process, accompanied by changes in the soft and hard composite strata. An improved mean denoising algorithm is proposed to mitigate the impact of significant differences in window values on denoising effectiveness. In each forward-sliding window process, the data within the window is sorted, and the maximum value  $y_{kmax}$  and the

minimum number  $y_{kmin}$  are excluded. The mean of the remaining data is then calculated to represent the measurement result. The formula for the improved algorithm is as follows:

$$\bar{y}_k = \frac{1}{2n-1} \left( \sum_{i=-n}^{2n+1} y_{k+i} - y_{kmax} - y_{kmin} \right), \tag{10}$$

In order to assess the denoising effect, this paper selects two evaluation metrics: Signal-to-Noise Ratio (SNR) and Peak Signal-to-Noise Ratio (PSNR). Signal-to-Noise Ratio (SNR): It evaluates signal strength and noise level ratio. Peak Signal-to-Noise Ratio (PSNR): It evaluates the ratio between the maximum possible signal and noise power. Excellent SNR and PSNR depend on the specific application scenarios and requirements. The metrics only need to reach a certain level, and in the context of denoising the same data, higher SNR and PSNR indicate better denoising effects. The formulas for calculating both metrics are as follows:

$$SNR = 10 * \lg \left( \frac{\sum p^2}{\sum (p - q)^2} \right), \tag{11}$$

$$PSNR = 10 * \lg \left( \frac{p_{max}^2}{\frac{1}{n} * \sum (p - q)^2} \right), \tag{12}$$

where  $p$  represents the original data,  $q$  is the filtered data, and  $p_{max}$  denotes the maximum signal value. In the engineering calculations, considering that the data processing is conducted in a ring-by-ring manner, with varying lengths of the excavation parameter sequences for each ring, a sliding window of size 11 is chosen after multiple tests. Here, the denoising effect after handling outliers is presented. Table 2 compares the evaluation metric values before and after algorithm improvement, while Figure 6 illustrates the denoising effect achieved via the improved mean filtering algorithm.

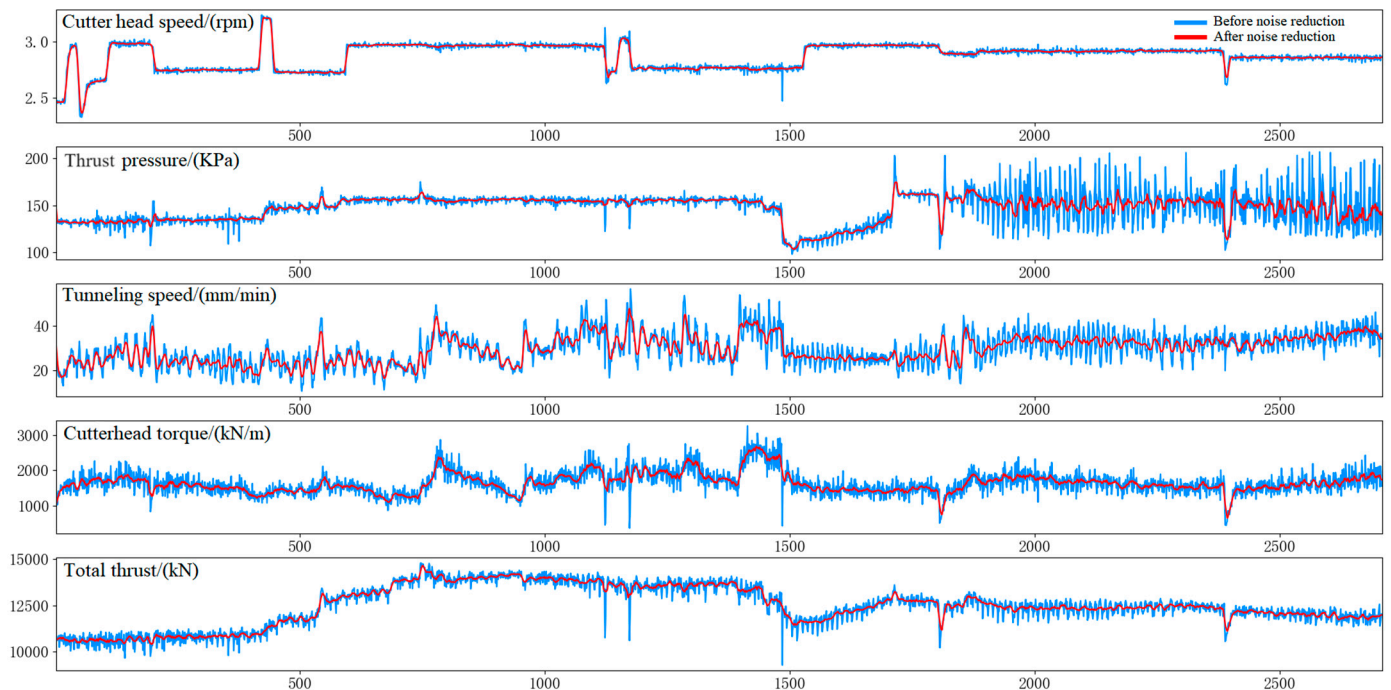


Figure 6. Comparison of variation amplitude of data values before and after denoising.

**Table 2.** Prediction effect ranking analysis.

	Cutterhead Speed		Thrust Pressure		Tunneling Speed		Cutterhead Torque		Total Thrust	
	SNR	PSNR	SNR	PSNR	SNR	PSNR	SNR	PSNR	SNR	PSNR
Mean filter algorithm	42.10	43.14	22.57	25.49	17.72	23.14	18.96	24.78	33.19	34.65
Improved mean filter algorithm	42.68	43.72	22.58	25.49	17.84	23.27	18.96	24.79	33.19	34.65

The improved algorithm enhances the denoising effect on the cutterhead rotation and tunneling speed. The SNR for the cutterhead rotation and tunneling speed increased by 0.58 and 0.12, respectively, while the PSNR increased by 0.58 and 0.13, respectively. The SNR for tunneling speed and cutterhead torque is below 20, indicating a relatively low signal-to-noise ratio. This suggests the presence of considerable noise in the data signal, likely influenced by changes in the geological conditions and factors such as vibrations and impacts in the mechanical system. This underscores the necessity of data denoising.

#### 4.2.4. Data Normalization

When data features have significantly different magnitudes and exhibit a wide range of values, normalization methods are commonly employed to balance the importance of different features in prediction, thereby enhancing prediction accuracy. Two commonly used normalization methods are min–max normalization and Z-score normalization. The computation formulas for these two methods are as follows:

$$x^* = \frac{x - x_{\min}}{x_{\max} - x_{\min}}, \quad (13)$$

$$x^* = \frac{x - x_{\text{mean}}}{\delta}, \quad (14)$$

As shown in Formula (13), the max–min normalization method linearly transforms the original data to the range [0, 1], where  $x_{\max}$  and  $x_{\min}$  represent the maximum and minimum values of the column data, respectively. This method is suitable for data distributions with clear boundaries, mainly when scaling the data to a fixed range is necessary. Formula (14) demonstrated that the z-score normalization method normalizes the original data set to have a mean of 0 and a standard deviation of 1. Here,  $x_{\text{mean}}$  represents the mean of all sample data, and  $\delta$  represents the standard deviation of all sample data. This method is applicable when data distribution lacks clear boundaries and is suitable for cases where comparing data from different features on the same scale is necessary. In the subsequent sections of this paper, different normalization methods will be applied to the training set, and the same method will then be used to normalize the validation set. The most suitable normalization method will be determined by comparing the model's predictive accuracy on the test set. It is important to note that considering the potential differences in data distribution between the training and validation sets, separate normalization processes for the training and validation sets are necessary to achieve optimal predictive performance.

#### 4.3. Evaluation Index

In the predictive model, various evaluation metrics are employed, each emphasizing different aspects, and relying on a single metric lacks comprehensiveness in model assessment. To evaluate the model's performance, three distinct evaluation metrics are employed: Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE), defined as follows:

$$MAE = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i|, \quad (15)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2}, \quad (16)$$

$$MAPE = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{\hat{y}_i - y_i}{y_i} \right|, \quad (17)$$

In the formula,  $y_i$  is the real value and  $\hat{y}_i$  is the predicted value. MAE is a commonly used metric for assessing the accuracy of predictive models, reflecting the degree of closeness between actual and predicted values. Similarly, RMSE is employed to gauge predictive model accuracy. In contrast to MAE, RMSE imposes a higher penalty by squaring the errors, making it more sensitive to significant errors and outliers. MAPE is highly sensitive to extreme values, limiting its ability to handle outliers effectively. However, it expresses errors as a percentage, providing interpretability and comparability.

## 5. Comparative Analysis of Forecast Results

### 5.1. Model Parameter Presets

The research process processed 1–680 tunnel boring machine (TBM) shield tunneling parameters. Due to the enormous volume of raw data and variations in data quality, which could impact the model's predictive performance, samples were selected from different tunneling modes. For the TBM mode, the training set comprised rings 120 to 138, with the prediction segment spanning rings 139 to 146. Under the Earth Pressure Balance (EPB) mode, the training set included rings 435 to 444, and the prediction segment covered rings 445 to 450. Approximately 310,000 data points for each parameter—tunneling speed, total thrust, cutterhead torque, tunneling pressure, and cutterhead speed—were selected, totaling around 1.55 million data points. The constructed model is a real-time prediction model with a rectified linear unit (ReLU) as the activation function. The time step was set to 20, utilizing the past 20 time steps of data as input to predict future data. The number of epochs was set to 200. Considering that optimization algorithms can optimize multiple hyperparameters in the LSTM model, such as the number of LSTM layers, the number of LSTM units in hidden layers, learning rate, batch size, etc. Given the limitations of existing computer performance, it is advisable to choose critical hyperparameters. In line with previous research experience, this study selected two commonly optimized hyperparameters before including optimization algorithms: the number of LSTM hidden layer neurons and the learning rate. Prior to incorporating optimization algorithms, preliminary tuning was conducted for other hyperparameters. The pre-training hyperparameter settings are presented in Table 3, with 30 combinations tested during the preliminary tuning, and the relative errors are compared in Figure 7.

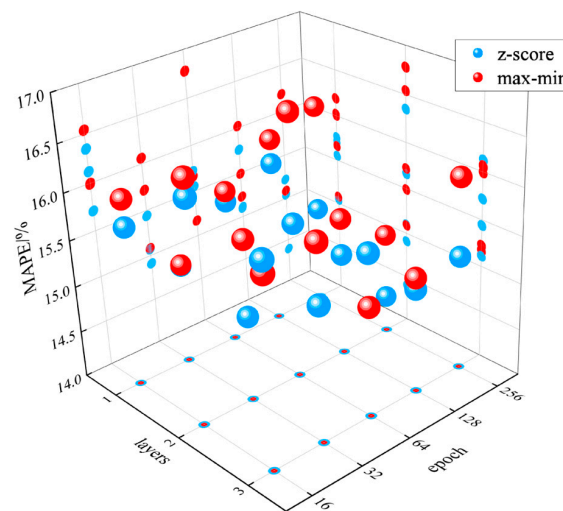


Figure 7. Comparison of relative errors in pre-training.

**Table 3.** Pre-training setup.

Name	Optimize Parameter Range
Normalization method	max-min normalization, z-score normalization
Batch size	16, 32, 64, 128, 256
Number of LSTM layers	1, 2, 3

After multiple pre-tuning experiments, the minimum average percentage error was 14.4%. Regarding the model’s accuracy, using the z-score normalization method for data processing, setting the number of LSTM layers to 2, and configuring the batch size to 256 were more effective in training the model. Considering the dimensional differences in the model’s input data, which could influence the model’s output, the z-score normalization method was employed to transform data with different dimensions into a unified scale. This ensures relative balance in the importance of these data during model computations, mitigating interference caused by dimensional disparities. The network structure becomes more complex when the number of LSTM layers is excessive. Although increasing the number of layers can enhance model performance and help capture deeper features and relationships, this improvement is dynamic. With the increased model complexity, the training process may face the risk of gradient explosions.

Additionally, computational resources and training time would also increase. After a series of pre-training experiments, it was determined that a two-layer LSTM configuration is optimal for the model. Furthermore, setting the batch size to 256 improves the model’s convergence speed and generalization ability. A larger batch size can simultaneously process more data, reducing noise and fluctuations. During training, using a batch size that is too small may lead to overfitting of the model to the training data, resulting in inadequate generalization of new data.

5.2. Excavation Rate Prediction Analysis

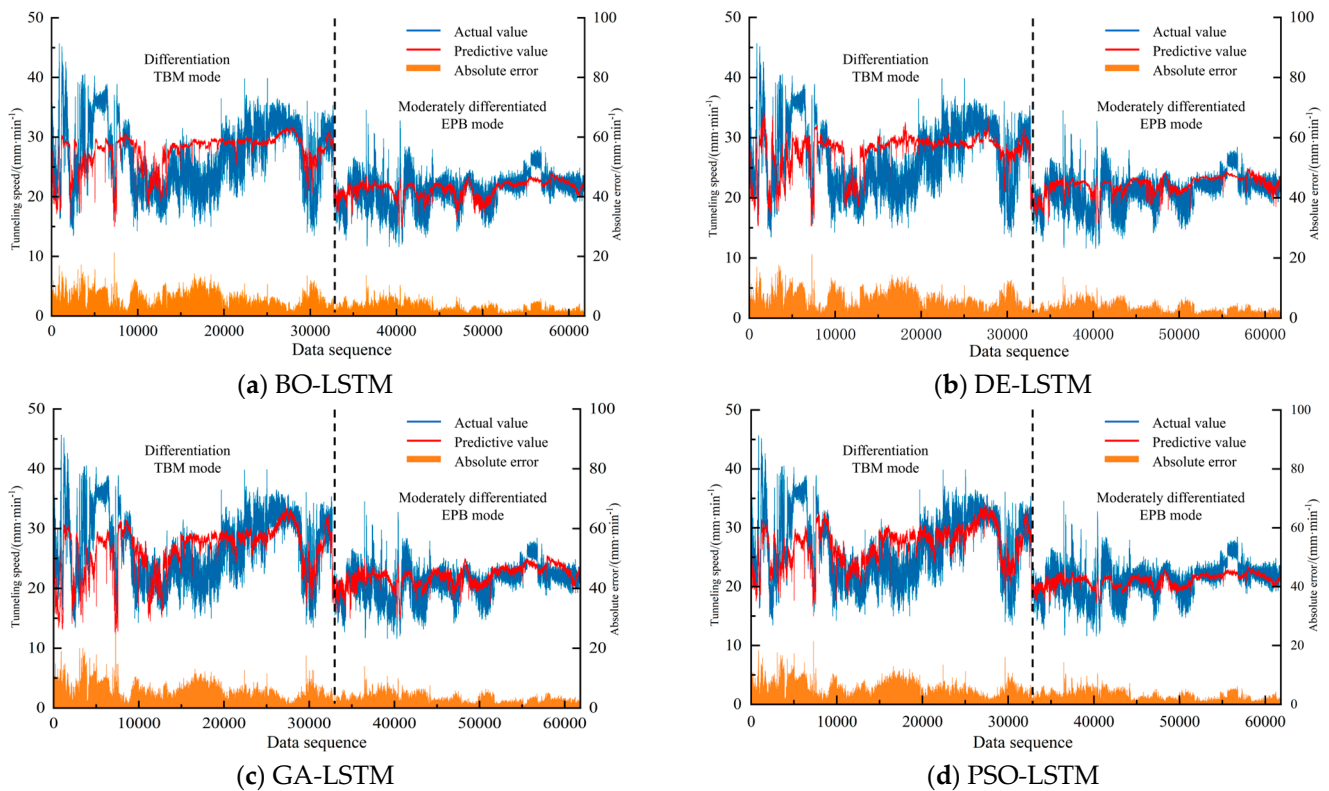
The four optimization algorithms set the optimization range for the number of neurons in the LSTM hidden layer between 10 and 100, and the optimization range for the learning rate is set between 0.001 and 0.1. At the same time, other parameters use the model’s default values. Figure 8 illustrates the predictive results of the LSTM model on the test set under different optimization algorithms. In regression models, different evaluation metrics focus on different aspects. Evaluating with a single metric may overlook other important factors, so a comprehensive evaluation method using multiple metrics is necessary. Zorlu et al. [29] proposed the ranking method as a commonly used multi-metric comprehensive evaluation method in 2008. This method involves ranking *N* models under the same evaluation metric, *i*. Then, the rankings of *m* different evaluation metrics for the same model are summed to obtain the multi-metric ranking for that model, as shown in Table 4.

**Table 4.** Prediction effect ranking analysis.

Predictive Model	Optimization Results		Training Time (s)	MAE		RMSE		MAPE		Overall Ranking	Rank Sort
	Units	Rate		EPB	TBM	EPB	TBM	EPB	TBM		
BO-LSTM	32	0.0094	367(1)	1.63(2)	3.51(1)	2.17(2)	4.49(1)	8.0%(1)	13.8%(1)	9	1
DE-LSTM	36	0.0085	800(3)	1.62(1)	3.93(3)	2.16(1)	4.88(3)	8.0%(1)	15.9%(3)	15	2
GA-LSTM	23	0.0086	1011(4)	1.82(4)	4.22(4)	2.38(4)	5.21(4)	9.3%(3)	16.9%(4)	27	4
PSO-LSTM	19	0.0150	396(2)	1.74(3)	3.77(2)	2.24(3)	4.83(2)	8.7%(2)	14.6%(2)	16	3

Note: ( ) is a single-column ranking.





**Figure 8.** Prediction results of different LSTM models for the test set.

After comparing the performance of different optimization models on the test set, the conclusion can be drawn that, regarding the optimization of model hyperparameters, the number of neurons in the LSTM hidden layer or the learning rate should be manageable. Many neurons in the LSTM hidden layer can lead to an overly complex model, making it prone to overfitting. Conversely, more neurons are needed to ensure the model captures complex patterns and relationships, resulting in an insufficient utilization of information from the input data. A learning rate that is too small may slow down network training, requiring a longer time to converge, while a learning rate that is too large may lead to unstable training, skipping the optimal points and preventing convergence. Therefore, it is necessary to use hyperparameter optimization algorithms for multiple training iterations to achieve optimal predictive performance.

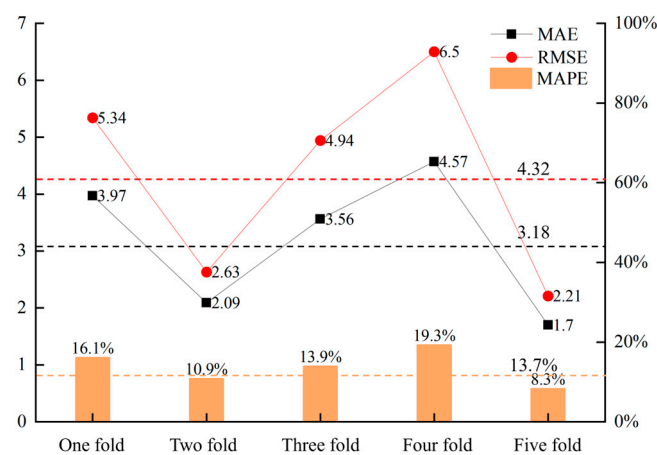
Examining the prediction curves for specific tunneling ring numbers reveals that the trends in the predicted data under both tunnel boring machine (TBM) and earth pressure balance (EPB) shield modes generally align with the actual excavation parameter curves. Under the TBM mode, the predicted curve for the weathered zone aligns well with the actual values, demonstrating overall good predictive performance. However, there is significant fluctuation in the local prediction segment from ring 139 to ring 141 due to unfavorable geological conditions characterized by complex rock changes in a poor geological area. This section underwent pre-reinforcement during construction, contributing to relatively poorer predictive results than other segments. In the EPB mode, the predicted curve for the composite soft and hard layer is smoother, reflecting stable changes in tunneling speed due to the softer rock characteristics, resulting in a closer match between predicted and actual values with more minor relative errors. Based on the overall ranking, BO-LSTM performs the best, with PSO-LSTM and DE-LSTM showing similar performance, while GA-LSTM performs the least favorably.

Regarding predictive accuracy, DE-LSTM and BO-LSTM are suitable for predicting relatively stable curve patterns. However, DE-LSTM performs poorly in predicting curves with significant fluctuations and suffers from the drawback of slow operation speed. Overall, BO-LSTM is a preferable model that can be applied to predict different geological

environments. In EPB mode, the Mean Absolute Percentage Error (MAPE) prediction result is 8%, while in TBM mode, the MAPE prediction result is 13.8%. PSO-LSTM exhibits good overall performance, providing accurate predictions with a shorter runtime, and can be considered an alternative prediction model.

### 5.3. Multi-Algorithm Optimization Model Prediction Analysis

The models above exhibit favorable predictive performance on a fixed test set, but this does not necessarily imply the same performance on the global data set. In order to enhance the model's generalization capability, dropout algorithms and five-fold time series cross-validation are introduced to the BO-LSTM model, proposing a multi-algorithm-optimized tunneling speed prediction model. The validation set is partitioned based on the temporal characteristics of the overall data set, and the evaluation metric results are depicted in Figure 9.



**Figure 9.** Time series cross-validation results.

In the above pictures, the dotted lines of different colors are the reference lines of the average value of the evaluation index. Via five rounds of time series cross-validation, the model demonstrates strong generalization capabilities, with a mean absolute error (MAE) of 3.18, root mean square error (RMSE) of 4.32, and mean absolute percentage error (MAPE) of 13.7%. In relatively stable geological layers, the MAPE prediction result is 8.3%. Even in challenging geological conditions where the excavation process experiences significant discrete fluctuations, the model maintains an 80% accuracy in predicting tunneling speed.

## 6. Conclusions

This paper relies on the Shenzhen Metro Line 13 Left Tunnel Project from Liuxiandong Station to Baimang Station. It utilizes geological parameters and real-time monitoring of excavation parameters in the complex strata to establish a prediction model for the excavation rate of EPB/TBM dual-mode tunnel boring machines (TBMs) based on multi-algorithm optimization and recurrent neural networks. The accuracy of the prediction results is analyzed, and the main conclusions are as follows:

- (1) The combined impact of the dual-mode TBM excavation mode, TBM operating state parameters, and geological parameter variations was considered in establishing the tunnel boring machine (TBM) excavation data set. The model is more interpretable under actual working conditions, supporting the planning and control of TBM construction.
- (2) Isolation Forest and an improved mean filtering algorithm were applied to handle TBM operating state parameters. This reduced the spatial variability of excavation parameters, making it easier for machine learning algorithms to learn the patterns of feature variations. The data set processed via denoising allows the construction of a more accurate prediction model for TBM tunneling speed.

- (3) A prediction model for TBM tunneling speed was established using multi-algorithm optimization and recurrent neural networks. The model integrates Bayesian optimization algorithms, dropout algorithms, and time-series cross-validation, demonstrating strong generalization capabilities and operational efficiency. The lowest MAPE prediction result is 8.3%, with an average MAPE prediction result below 15%.

The established model primarily aims to explore the real-time operational patterns of tunneling speed in the dual-mode TBM excavation process under unsupervised training models and provide a foundational model for intelligent decision control. In practice, the data in the tunneling process is input into the model in real time, the model is trained, and the results are output in real time, which plays a guiding role in the field. In this study, the mechanical properties of rock have not been considered, and the model is not refined enough. In order to obtain a refined model that is more generalized and more in line with engineering practice, future research will consider more input of tunneling parameters and rock mechanics parameters.

**Author Contributions:** Conceptualization, T.Y. and T.W.; methodology, T.Y., X.H., G.S. and B.L.; software, T.W.; validation, B.L., S.L., G.S., H.S. and T.W.; formal analysis, X.P. and X.H.; investigation, B.L., X.P., H.S. and S.L.; resources, T.W.; data curation, T.Y., H.S. and S.L.; writing—original draft preparation, T.Y., G.S., T.W. and X.H.; writing—review and editing, T.W.; visualization, B.L., G.S. and X.P.; supervision, X.H. and B.L.; project administration, B.L., G.S., H.S. and S.L.; funding acquisition, X.H. and S.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** It was supported by the National Natural Science Foundation of China Regional Joint Key Project (Grant No. U22A20234), Wuhan City Knowledge Innovation Special Project (Grant No. 2022010801010162), the National Natural Science Foundation of China (Grant No. 52074258), Hubei Provincial Key Research and Development Program (Grant No. 2021BCA133), Hubei Provincial Natural Science Foundation Outstanding Youth Project (Grant No. 2022CFA084) and China State Construction Engineering Corporation Science and Technology Research and Development Program Support (Grant No. CSCEC-2019-Z-19).

**Data Availability Statement:** Data are contained within the article.

**Conflicts of Interest:** Author Hongbing Shi was employed by China Construction Civil Infrastructure Corp., Ltd. Author Shaoran Liu was employed by China Construction South Investment Co., Ltd. Author Guangzu Sheng was employed by Wuhan Urban Construction Group Construction Management Co., Ltd. The funding sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results.

## References

1. Song, T.; Liu, C.; Chen, F.; Xu, Y.; Wang, S.; He, C. Research on the Selection and Tunneling Adaptability of Dual-mode Shield/TBM in Metro Tunnel. *J. Railway Eng. Soc.* **2022**, *39*, 17.
2. Editorial Department of China Highway and Transport. Review on China's Traffic Tunnel Engineering Research: 2022. *China Highw. Transp.* **2022**, *35*, 1–40. [[CrossRef](#)]
3. Chen, F.; Huang, Z.; He, C.; Meng, Q.; Lai, X.; Liu, C.; Wang, S. Selection of reasonable mode conversion point for earth pressure/slurry dual-mode shield in round gravel-mudstone composite strata. *China Civ. Eng. J.* **2021**, *54*, 48–57. [[CrossRef](#)]
4. He, C.; Chen, F.; Huang, Z.; Meng, Q.; Liu, C.; Wang, S. Adaptability and Excavation Parameters of Compound Strata with Dual-Mode Shield Tunneling. *Chin. J. Geotech. Eng.* **2021**, *43*, 43–52.
5. Xie, X.; Yang, C.; Wang, Q.; Zeng, L.; Hou, J.; Zhou, B. Settlement Analysis and Control Research on Shield Tunnel Crossing the Yangtze River Embankment in Nanjing and Yan Road Overpass. *Chin. J. Rock Mech. Eng.* **2021**, *40*, 3313–3322.
6. Shen, X.; Yuan, D.J.; Jin, D.L. Influence of shield attitude change on shield–soil interaction. *Appl. Sci.* **2019**, *9*, 1812. [[CrossRef](#)]
7. Jin, D.L.; Shen, X.; Yuan, D.J. Theoretical analysis of three-dimensional ground displacements induced by shield tunneling. *Appl. Math. Model.* **2020**, *79*, 85–105.
8. Chen, C.; Shi, P.; Jia, P.; Dong, M. Correlation analysis of shield driving parameters and structural deformation prediction based on MK-LSTM algorithm. *J. Jilin Univ. (Eng. Technol. Ed.)* **2023**, 1–10. [[CrossRef](#)]
9. Zhao, B.; Zhou, J.; Tan, Z. Variation of shield boring parameters and correlation analysis in mixed ground. *China Civ. Eng. J.* **2017**, *50*, 140–144.
10. Qin, R.; Li, X. Influence of shield driving speed on ground surface settlement in clay layers. *China Civ. Eng. J.* **2020**, *53*, 1–6.

11. Zhao, J.; Gong, Q.M.; Eisensten, Z. Tunnelling through a frequently changing and mixed ground: A case history in Singapore. *Tunn. Undergr. Space Technol.* **2007**, *22*, 388–400. [[CrossRef](#)]
12. Farrokh, E.; Rostami, J. Correlation of tunnel convergence with TBM operational parameters and chip size in the Ghomroud tunnel, Iran. *Tunn. Undergr. Space Technol.* **2008**, *23*, 700–710. [[CrossRef](#)]
13. Lüth, S.; Giese, R.; Otto, P.; Krüger, K.; Mielitz, S.; Bohlen, T.; Dickmann, T. Seismic investigations of the Piora Basin using S-wave conversions at the tunnel face of the Piora adit (Gotthard Base Tunnel). *Int. J. Rock Mech. Min. Sci.* **2008**, *45*, 86–93. [[CrossRef](#)]
14. Sapigni, M.; Berti, M.; Bethaz, E.; Busillo, A.; Cardone, G. TBM performance estimation using rock mass classifications. *Int. J. Rock Mech. Min. Sci.* **2002**, *39*, 771–788. [[CrossRef](#)]
15. Kahraman, S.; Bilgin, N.; Feridunoglu, C. Dominant rock properties affecting the penetration rate of percussive drills. *Int. J. Rock Mech. Min. Sci.* **2003**, *40*, 711–723. [[CrossRef](#)]
16. Hassanpour, J.; Rostami, J.; Khamehchiyan, M.; Bruland, A. Developing new equations for TBM performance prediction in carbonate-argillaceous rocks: A case history of Nowsod water conveyance tunn. *Geomech. Geoengin. Int. J.* **2009**, *4*, 287–297. [[CrossRef](#)]
17. Wang, H.; Fu, D. Mathematical and physical model and relationship analysis among various parameters in earth pressure balance shield tunneling. *China Civ. Eng. J.* **2006**, *39*, 86–90.
18. Zhang, Z.; Li, T.; Han, A.; Su, M.; Huang, X. Prediction model of shield tunneling speed and cutterhead torque in complex strata and its study on strata adaptability. *Tunn. Constr.* **2016**, *36*, 1449–1455.
19. Xu, Q. Study on the Simulated Model Test of Shield Machine' Working Parameters Applicable to Different Stratums and Its Theoretical Investigation. Ph.D. Thesis, Tongji University, Shanghai, China, 2006.
20. Xu, H.; Zhou, J.; Asteris, P.G.; Jahed Armaghani, D.; Tahir, M.M. Supervised machine learning techniques to the prediction of tunnel boring machine penetration rate. *Appl. Sci.* **2019**, *9*, 3715. [[CrossRef](#)]
21. Zhang, Z.; Li, X.; Ji, J. Prediction model of TBM excavation parameters based on LS-SVM. *J. Hohai Univ. (Nat. Sci.)* **2021**, *49*, 373–379.
22. Li, C.; Li, T.; Li, Z.; Zhan, J. Prediction and Analysis of Shield Tunneling Parameters in Compound Strata Based on BP Neural Network. *China Civ. Eng. J.* **2017**, *50*, 145–150.
23. Hou, S.; Liu, Y.; Zhang, K. Prediction of TBM Excavation Parameters Based on IPSO-BP Hybrid Model. *Chin. J. Rock Mech. Eng.* **2020**, *39*, 1648–1657.
24. Qiu, D.; Fu, K.; Xue, Y.; Li, Z.; Li, G.; Kong, F. LSTM time-series prediction model for TBM tunneling parameters of deep-buried tunnels and application research. *J. Cent. S. Univ. (Sci. Technol.)* **2021**, *52*, 2646–2660.
25. GB 50307-2012; Code for Geotechnical Investigations of Urban Rail Transit. China Planning Press: Beijing, China, 2012.
26. Wang, S.; Wang, Y.; Li, X.; Liu, L.; Yin, T. Study of Standardized Pre-processing Method of TBM Tunneling Data. *Mod. Tunn. Technol.* **2022**, *59*, 38–44.
27. Liu, F.T.; Ting, K.M.; Zhou, Z.H. Isolation-based anomaly detection. In Proceedings of the 2008 Eighth IEEE International Conference on Data Mining, Pisa, Italy, 15–19 December 2008; pp. 413–422.
28. Xiao, H.H.; Yang, W.K.; Hu, J.; Zhang, Y.P.; Jing, L.J.; Chen, Z.Y. Significance and methodology: Preprocessing the big data for machine learning on TBM performance. *Undergr. Space* **2022**, *7*, 680–701. [[CrossRef](#)]
29. Zorlu, K.; Gokceoglu, C.; Ocakoglu, F.; Nefeslioglu, H.A.; Acikalin, S.J.E.G. Prediction of uniaxial compressive strength of sandstones using petrography—Based models. *Eng. Geol.* **2008**, *96*, 141–158. [[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.