Journal of Rock Mechanics and Geotechnical Engineering xxx (xxxx) xxx

Contents lists available at ScienceDirect

Journal of Rock Mechanics and Geotechnical Engineering

journal homepage: www.jrmge.cn

Full Length Article

An explainable deep learning approach to enhance the prediction of shield tunnel deviation

Jiajie Zhen [a], Fengwen Lai [b], Ming Huang [b,*], Junjie Zheng [c], Jim S. Shiau [d], Ping Wang [e], Jinhuo Zheng [f]

[a] College of Transportation and Civil Engineering, Fujian Agriculture and Forestry University, Fuzhou, 350108, China
[b] College of Civil Engineering, Fuzhou University, Fuzhou, 350108, China
[c] School of Civil Engineering, Wuhan University, Wuhan, 430072, China
[d] School of Engineering, University of Southern Queensland, Toowoomba, QLD, 4350, Australia
[e] China Communications Construction First Highway Engineering Bureau Xiamen Co., Ltd., Xiamen, 361021, China
[f] Fujian Provincial Institute of Architectural Design and Research Co., Ltd, China

ARTICLE INFO

Article history:
Received 7 July 2025
Received in revised form
5 November 2025
Accepted 14 November 2025
Available online xxx

Keywords:
Shield tunnel position deviation
Machine learning
Explainable AI
Deep learning important features

ABSTRACT

Although machine learning models have achieved high enough accuracy in predicting shield position deviations, their "black box" nature makes the prediction mechanisms and decision-making processes opaque, leading to weaker explanations and practicability. This study introduces a novel explainable deep learning framework comprising the Informer model with enhanced attention mechanisms (EAMInfor) and deep learning important features (DeepLIFT), aimed at improving the prediction accuracy of shield position deviations and providing interpretability for predictive results. The EAMInfor model attempts to integrate channel attention, spatial attention, and simple attention modules to improve the Informer model's performance. The framework is tested with the four different geological conditions datasets generated from the Xiamen metro line 3, China. Results show that the EAMInfor model outperforms the traditional Informer and comparison models. The analysis with the DeepLIFT method indicates that the push thrust of push cylinder and the earth chamber pressure are the most significant features, while the stroke length of the push cylinder demonstrated lower importance. Furthermore, the variation trends in the significance of data points within input sequences exhibit substantial differences between single and composite strata. This framework not only improves predictive accuracy but also strengthens the credibility and reliability of the results.

1. Introduction

With the growing demand for reliability and safety, shield tunneling has become one of the preferred methods for constructing metros (Huang et al., 2023; Lai et al., 2023, 2024; Zhou et al., 2023). In this context, controlling the position of shield machines has received increasing attention, particularly due to frequent encounters with the position's deviations from their designated tunnel axes (DTA) (Cao et al., 2024; Xu et al., 2024). Currently, shield machines are primarily operated manually, with drivers adjusting based on observed deviations and their experience (Shen et al., 2023; Zhou et al., 2023). Predicting the shield's position and attitude in advance could provide drivers with timely data-driven guidance for more precise and effective control. Unchecked deviations can significantly impact tunnel quality, leading to issues such as misaligned tunnel linings, water infiltration, or even the destabilization and overturning of the shield machine (Zhou et al., 2019; Xu et al., 2023).

With the rapid advancement of artificial intelligence and big data technologies (Lai et al., 2023, 2025; Duan et al., 2024a; Chen et al., 2025), machine learning (ML) models have been increasingly used to predict shield tunnel position deviations (Wu et al., 2021a; Dai et al., 2023; Zheng et al., 2024), assist in decision-making for boring parameters (Liu et al., 2021; Wang et al., 2023a; Lu et al., 2024), predict geologic condition (Fang et al., 2023; Duan et al., 2024b; Tan et al., 2025a), and assess surface settlement (Zhang

* Corresponding author.
E-mail address: huangming05@163.com (M. Huang).

https://doi.org/10.1016/j.jrmge.2025.11.002

et al., 2019, 2022; Wang et al., 2024). ML models such as random forest (RF) (Zhang et al., 2019; Shen et al., 2023), long short-term memory (LSTM) networks (Xu et al., 2023; Chen et al., 2024), and gated recurrent units (GRU) (Xiao et al., 2022; Wang et al., 2023b) have been successfully applied to predict shield position and attitude parameters with high accuracy. While these studies have achieved high enough accuracy in predicting shield tunneling position and attitude, the "black box" nature of these models makes it difficult to understand the basis for their decisions, further hindering their credibility and practical applicability.

Recent research (Chen et al., 2022; Ortigossa et al., 2024) has revealed that even state-of-the-art models can exhibit logical inconsistencies. When predictions deviate or exhibit errors, the lack of interpretability makes it challenging for engineers to trace the issue back to specific input features or parameter settings (Zhou et al., 2023). This complicates troubleshooting, delays resolution, and increases the risk of incorrect decisions. For instance, engineers may incorrectly attribute prediction errors to improper adjustments of construction parameters rather than to issues within the model itself. Such misjudgments can lead to misguided corrective actions and heightened operational risks. Furthermore, the lack of interpretability implies that if models rely on superficial or irrelevant features during training, such issues may remain undetected, thereby increasing potential risks in real-world applications. To gain wider acceptance, especially in high-risk environments where incorrect decisions can have severe consequences, models must provide clear explanations of their decision-making processes and outcomes (Chen and Fan, 2023; Lin et al., 2024).

Indeed, current ML studies in the field of shield tunneling have placed limited emphasis on model interpretability. Hu et al. (2022) utilized the XGBoost model to predict shield tunnel positional deviation, leveraging the model's interpretability to directly output importance scores of input features. However, the prediction accuracy of tree-based models in some complex shield tunneling projects is often lower than that of deep learning models, such as recurrent neural networks (Zhou et al., 2022; Xu et al., 2023). Fu et al. (2023) proposed a GCN-LSTM model to predict vertical and horizontal deviations of shield machines and analyzed feature importance scores using the Shapley additive explanations (SHAP) method (Lundberg and Lee, 2017). However, the SHAP method is not directly interpretable for the GCN-LSTM model and requires a separate setup for SHAP analysis. This limitation means the internal decision-making mechanisms of complex deep learning models cannot be fully explored using standard SHAP methods.

Explainable Artificial Intelligence (XAI) is a promising technology that improves model interpretability by using specific algorithms to relate input features to output information (Linardatos et al., 2021). XAI method can be divided into intrinsic methods and post-hoc methods. Intrinsic interpretability refers to models that inherently possess interpretability during their design and training processes, such as linear or tree-based models. While deep learning models, particularly those of the Transformer class, offer some degree of intrinsic interpretability through mechanisms like attention, time encoders, and decoders, these are not always intuitive and can be difficult to fully comprehend. Therefore, post-hoc interpretability methods, such as deep learning important features (DeepLIFT), are often employed to explain a model's output by calculating the contribution of each input feature or data point in the input sequence to the model's decisions (Shrikumar et al., 2017). Notably, the accuracy of a model's predictions is crucial for establishing the credibility of its interpretability. If the model's predictions are erroneous, the interpretability of its results may also be compromised, potentially failing to accurately represent the relationship between features and predictions. In other words, the interpretability of a model is inherently dependent on its capacity to accurately capture the underlying patterns in the data; otherwise, the resulting interpretations may lack meaningful significance.

This study introduces an innovative Informer model (EAMInfor) that integrates the channel attention mechanism, spatial attention mechanism, and simple attention mechanism (SimAM) to improve accuracy in predicting deviations in shield tunneling position. Subsequently, the contribution scores of input features and individual data points in the input sequence are determined using the DeepLIFT method. Finally, the Interpretability result of the DeepLIFT method is further validated through experiments with various input feature combinations and the SHAP method. The proposed method improves the accuracy of shield tunneling position deviation prediction and strengthens the credibility and reliability of the deep learning model for practical applications.

2. Methodologies

2.1. Informer model with enhanced attention mechanisms (EAMInfor)

Complex temporal and spatial correlations exist among shield boring parameters and shield tunneling position deviation parameters. The shield tunneling monitoring dataset exhibits temporal dependencies, where past states influence future outcomes. Highly dynamic features, such as cutterhead torque and earth chamber pressure, fluctuate frequently, whereas low-dynamic features, such as the screw conveyor rotation speed and screw conveyor pressure, change more gradually. These characteristics necessitate the development of models capable of capturing both temporal and spatial correlations, while effectively extracting global trends and local patterns.

The conventional Informer model has effectively captured long-term dependencies along the temporal dimension using the ProbSparse self-attention mechanism. However, it struggles to effectively learn the complex interactions among multiple features. To address this limitation, a channel attention mechanism is introduced to dynamically assign importance weights to different features, enabling the model to prioritize the most critical features for accurate predictions. Similarly, the spatial attention mechanism helps the model capture interactions among multivariate features.

Furthermore, complex interactions exist among the various parameters within the shield tunneling monitoring dataset, requiring the ML model to effectively capture spatial correlations between input features. SimAM effectively captures global dependencies across the temporal and feature dimensions of the input sequence, making it particularly well-suited for modeling feature interactions and complex dependencies over extended periods. Notably, SimAM is a lightweight attention mechanism module that significantly enhances performance while introducing minimal additional parameters and computational overhead, making it an efficient and practical solution for real-world applications.

Therefore, we propose an Informer model based on the channel attention mechanism, the spatial attention mechanism, and the SimAM attention mechanism. Moreover, an additive residual connection (He et al., 2016) is employed to tackle challenges such as gradient vanishing that arise as model depth increases, ensuring stable and consistent performance improvements. The model structure is shown in Fig. 1. By integrating these attention mechanisms with additive residual connections, the proposed model effectively leverages the strengths of channel, spatial, and

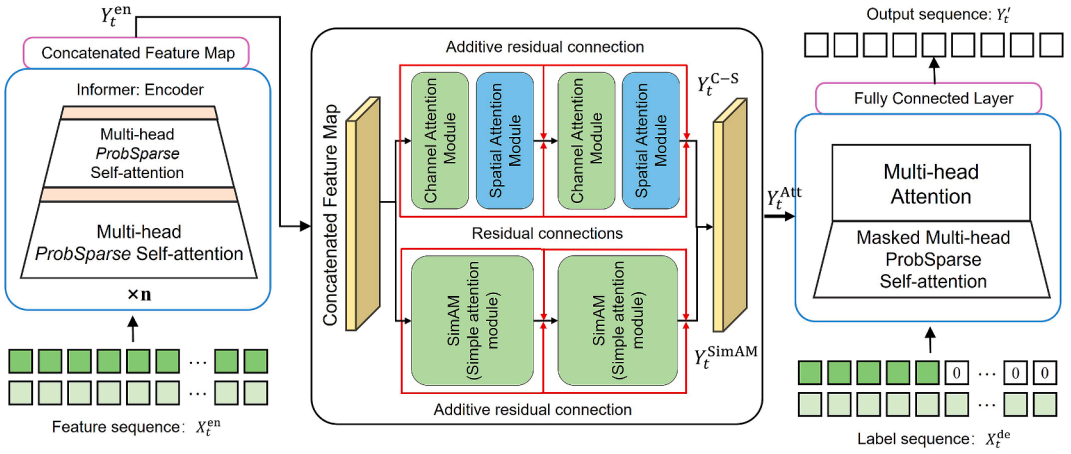

Fig. 1. Structure of EAMInfor model. The red line represents an additive residual connection.

spatiotemporal relationships. This collaborative design significantly enhances the model's capability to represent complex shield tunneling time-series data, enabling more accurate and robust predictions.

The input X_t^{en} is processed through encoding and distillation layers. The ProbSparse self-attention mechanism reduces time complexity and memory usage by focusing on higher-scoring dot products and ignoring lower-scoring ones:

$$\mathrm{MulAtt}(Q,K,V)=\mathrm{softmax}\left(\frac{\overline{Q}K^T}{\sqrt{d}}V\right) \tag{1}$$

where Q, K, and V represent the query matrix, key matrix, and value matrix, respectively; $\overline{Q}$ denotes the matrix obtained after scalarization of the Q matrix; $\sqrt{d}$ denotes a scaling factor introduced to prevent gradient vanishing; Softmax represents the activation function; and MulAtt is a ProbSparse self-attention mechanism.

The coding layer selectively extracts key features through distillation operations and generates self-attentive feature maps in subsequent layers. The formula for this extraction process from layer j to layer $j+1$ is as follows:

$$X_{j+1}^t=\mathrm{MaxPool}\left(\mathrm{ELU}\left(\mathrm{Convld}\left(\left[X_j^t\right]_{\mathrm{AB}}\right)\right)\right) \tag{2}$$

where $[\cdot]_{\mathrm{AB}}$ represents ProbSparse self-attention mechanism, MaxPool denotes maximum pooling operation. Convld represents a function that performs normalized convolution operations along the time series using the ELU activation function.

The output X_t^{en} through the encoder layer is Y_t^{en}, then fed into the channel attention mechanism, spatial attention mechanism, and SimAM (Yang et al., 2021) layers, respectively.

Y_t^{en} passes through both the channel attention module and the spatial attention module. The dimensions of the layer output Y_t^{en-CS} are identical to those of the input. In the SimAM layer, the self-similarity of each location in the feature map Y_t^{en} with other locations is calculated. The dimensions of the SimAM output $Y_t^{en-SimAM}$ are the same as the input dimensions.

The output of the multiple attention mechanisms layer Y_t^{en-Att} is

$$Y_t^{en-Att}=Y_t^{en-CS}\oplus Y_t^{en-SimAM} \tag{3}$$

The input time series X_t^{de} consists of two components: the historical input series X_t^{token}, which serves as a reference, and the predicted series X_t^0, which needs to mask future characterization factors:

$$X_t^{de}=\mathrm{Concat}\left(X_t^{token},X_t^0\right)\in R^{(L_{token}+L_y)d_{model}} \tag{4}$$

where $X_t^{token}\in\mathbb{R}^{L_{token}\times d_{model}}$ represents the start token, $X_t^0\in\mathbb{R}^{L_y\times d_{model}}$ denotes a placeholder for the target sequence.

Y_t^{en-Att} is fed into the multi-head attention mechanisms of the decoder layer. The multi-head attention mechanisms in the decoder layer differ from the ProbSparse self-attention mechanisms in the encoder layer. The calculation formula is as follows:

$$\mathrm{MulA}(\overline{Q},\overline{K},\overline{V})=\mathrm{Concat}\left(h_1^{de},h_2^{de},\ldots,h_i^{de}\right) \tag{5}$$

$$h_i^{de}=\mathrm{Attention}\left(QW_i^Q,KW_i^K,VW_i^V\right) \tag{6}$$

where W_i^Q, W_i^K, W_i^V represents the weight matrix for learning.

Finally, the high-dimensional feature Y_t^{de} is mapped to a relatively low-dimensional output sequence through a fully connected layer, followed by a ReLU activation function:

$$Y_t'=\mathrm{ReLU}\left(O_{\mathrm{wfc}}\times Y_t^{de}+b_{\mathrm{fc}}\right) \tag{7}$$

where O_{wfc} and b_{fc} represent learnable parameters, and Y_t' denotes the output of the EAMInfor model.

Notably, the additive residual connection (see the red line of Fig. 1) was integrated into the enhanced attention mechanisms module. The approach ensures the preservation and improvement of the learning performance of the model, even as its depth increases substantially. The output of the Informer encoder Y_t^{en} is fed into channel attention, spatial attention, and SimAM modules. Assuming that their processing functions are $f_{C-S}(x)$ and $f_{\mathrm{SimAM}}(x)$, the outputs of the improved attention mechanisms module that incorporates an additive residual connection are given as follows:

$$Y_t^{en-CS}=Y_t^{en}+f_{C-S}(x) \tag{8}$$

$$Y_t^{en-SimAM}=Y_t^{en}+f_{\mathrm{SimAM}}(x) \tag{9}$$

2.2. Deep learning important features (DeepLIFT) method

DeepLIFT (Shrikumar et al., 2017) is an additive feature attribution technique specifically designed for deep neural networks. DeepLIFT compares the activation of each neuron to its "reference activation" and assigns an importance score to each input based on this difference. The "reference activation" is determined using a user-defined reference input that represents an uninformative baseline value.

Let t denotes some target output neuron of interest, and let x_1, x_2, …, x_n represent some neurons in an intermediate layer or layer group. Define Δt as the difference from the reference, i.e., $\Delta t = t - t_0$. DeepLIFT assigns the contribution score $C_{\Delta x_i \Delta t}$ to Δ_{x_i} with the following formula:

$$\sum_{i=1}^{n} C_{\Delta x_i \Delta t} = \Delta t \tag{10}$$

For a given input neuron x (reference difference Δx) and target neuron t (reference difference), their contribution is computed by the multiplier $m\Delta x\Delta t$, which defines the contribution of Δx to Δt divided by Δx:

$$m\Delta x\Delta t = \frac{C\Delta x\Delta t}{\Delta x} \tag{11}$$

Given the value of $m\Delta x_i\Delta y_j$ and $m\Delta y_j\Delta t$, $m\Delta x_i\Delta t$ denotes defined as follows:

$$m_{\Delta x_i \Delta t} = \sum_j m_{\Delta x_i \Delta y_i} m_{\Delta y_j \Delta t} \tag{12}$$

DeepLIFT was selected in this study for two main reasons. First, it effectively handles discontinuities in the gradient of the Informer model by utilizing a difference from the reference approach. Second, it avoids the problem of model saturation, where using gradients would merely assign zero to features. Moreover, DeepLIFT can assign a non-zero score to these features. In addition, the feature ranking is generated through a single backpropagation pass through the network, which facilitates efficient explanation generation. This efficiency can help the shield driver by providing explanations alongside predictions.

3. Dataset creation

3.1. Background

The Xiamen metro line 3 features a double-bore, single-lane tunnel design, spanning a total length of 38.47 km. The study area focused on a segment from DK33 + 622 to DK33 + 732.4, primarily traversing silty powdery clay, residual gravelly clay, and fully weathered granite, with a burial depth ranging from 11.2 m to 12 m. The project employed an earth pressure balance (EPB) shield machine for tunneling, equipped with a cutterhead with a rotational speed of 0–3.7 rpm, a torque of 5631 kN m, a maximum thrust of 4086 T, and a peak propulsion speed of 80 mm/min.

Geological conditions at a site can significantly impact shield tunneling position deviations, and the distribution of tunneling parameters varies widely depending on these conditions. This variability is certain to affect the focus and performance of the ML model. To assess the impact of various geological conditions on shield tunnel position deviations, four datasets of different geological conditions were selected, such as residual gravelly clayey (RGC) soil, fully weathered granite mixed with silty clay (FWG-SC), fully weathered granite (FWG), and marine reclaimed sand (MRS), as detailed in Table 1.

3.2. Dataset preprocessing

The complex construction environment of underground spaces, combined with challenges in data transmission, often results in time-series data recorded by the shield tunneling machine's sensors containing missing values and outliers (Tan et al., 2025b). To accurately uncover the underlying correlations between input features and shield position deviation, it is essential for the dataset to effectively capture the dynamic spatiotemporal variation patterns of various parameters during each excavation step. Therefore, the dataset exported from the shield machine's onboard computer underwent essential preprocessing to ensure data quality and reliability. The data preprocessing workflow is illustrated as follows:

First, removal of abnormal excavation step data: Data from excavation steps involving multiple forced stoppages or other irregular operations were deleted. Second, removal of downtime phase data: Data recorded during segment assembly and maintenance phases were removed, as the advance speed (AS) and cutterhead rotation speed (CRS) were zero during these periods (Shen et al., 2023; Xu et al., 2023). Third, outlier handling: Outliers were identified using the density-based spatial clustering of applications with noise (DBSCAN) method. DBSCAN, a density-based clustering approach (Schubert et al., 2017), does not rely on assumptions about the shape of data distribution, making it well-suited for the nonlinear distribution of tunneling data. Fourth, missing value handling: Since abnormal excavation step data had already been removed in the earlier stages, the dataset contained only minor point or block missing data, with no long-interval continuous missing values. Missing values were imputed using forward-fill techniques, which fill missing data points with the most recent valid value (Seu et al., 2022). A visual representation of the data preprocessing steps is provided in Fig. 2.

3.3. Model input features and prediction targets

To thoroughly investigate how shield boring parameters impact the ML model's prediction of shield position deviations, 22 input features were selected based on relevant studies (Wang et al., 2019; Zhou et al., 2019, 2023; Xiao et al., 2021, 2022; Xu et al., 2023; Shen et al., 2023), including push thrust of push cylinder in group A (PT-A), push thrust of push cylinder in group B (PT-B), push thrust of push cylinder in group C (PT-C), push thrust of push cylinder in group D (PT-D), stroke length of push cylinder in group A (SPC-A), stroke length of push cylinder in group B (SPC-B), stroke length of push cylinder in group C (SPC-C), stroke length of push cylinder in group D (SPC-D), center left earth chamber pressure (CL-CP), upper left earth chamber pressure (UL-CP), lower left earth chamber pressure (LL-CP), lower right earth chamber pressure (LR-CP), center right earth chamber pressure (CR-CP), screw conveyor pressure (SCP), screw conveyor rotation speed (SCRS), screw conveyor torque (SCT), advance speed (AS), total thrust force (TTF), penetration (P), cutterhead torque (CT), cutterhead rotation speed (CRS), and cutterhead pressure (CP). The data distributions for each input feature are shown in Table 2.

The 22 input features are categorized into five sections: propulsion parameters (PT-A, PT-B, PT-C, PT-D, SPC-A, SPC-B, SPC-C, SPC-D, TTF), earth chamber pressure parameters (CL-CP, UL-CP, LL-CP, LR-CP, CR-CP), screw conveyor parameters (SCP, SCRS, SCT), cutterhead parameters (CT, CRS, CP), and advance parameters (AS, P). The EPB shield machine used for the Xiamen Metro Line 3 tunnel features a four-zone thrust design, where the push cylinders are divided into four groups (A, B, C, and D), as illustrated in Fig. 3a. The stroke length of the push cylinders (SPC-A, SPC-B, SPC-C, SPC-D) corresponds to the cylinder stroke of each group

Table 1
Datasets for four different geologic conditions.

Dataset	RGC dataset	FWG-SC dataset	FWG dataset	MRS dataset
Ring number	276–361	459–560	698–858	1159–1270
Number of rings	85	101	160	111
Stratum	Residual gravelly clayey soil (RGC)	Fully weathered granite mixed with silty clay (FWG-SC)	Fully weathered granite (FWG)	Marine reclaimed sand (MRS)

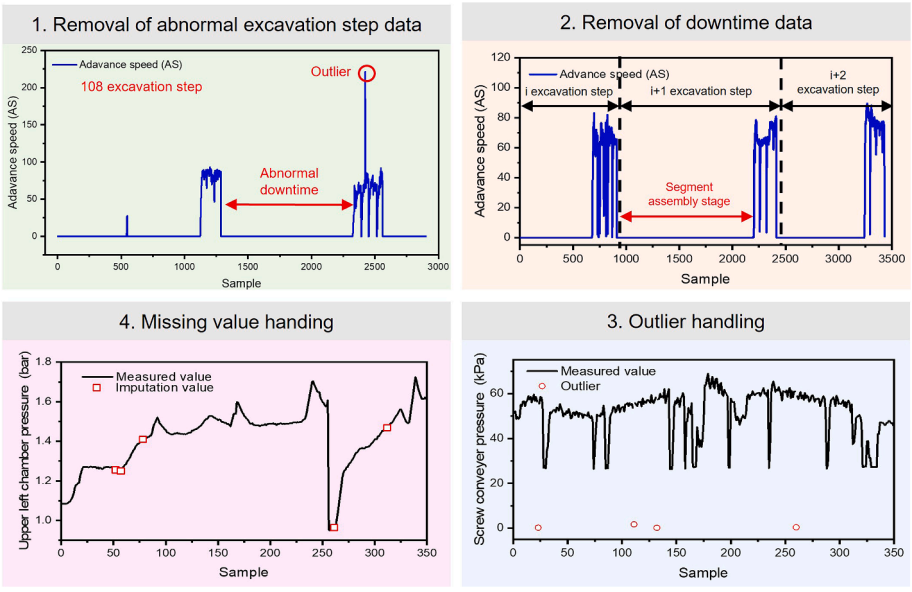

Fig. 2. Dataset preprocessing flowchart.

Table 2
Statistics of input features.

Parameter	Unit	Max	Min	Average	Median
Push thrust of push cylinder in group A (PT-A)	kN	248.26	29.07	108.72	104.05
Push thrust of push cylinder in group B (PT-B)	kN	298.47	48.05	173.78	179.95
Push thrust of push cylinder in group C (PT-C)	kN	196.26	13.02	87.27	82.17
Push thrust of push cylinder in group D (PT-D)	kN	181.23	31.21	91.33	83.64
Stroke length of push cylinder in group A (SPC-A)	mm	1737.88	645.89	1183.77	1181.63
Stroke length of push cylinder in group B (SPC-B)	mm	1737.32	643.78	1187.81	1186.03
Stroke length of push cylinder in group C (SPC-C)	mm	1703.87	673.53	1188.31	1188.81
Stroke length of push cylinder in group D (SPC-D)	mm	1731.37	680.50	1199.74	1199.55
Center left earth chamber pressure (CL-CP)	bar	3.03	0.99	1.81	1.69
Upper left earth chamber pressure (UL-CP)	bar	1.82	1.19	1.52	1.51
Lower left earth chamber pressure (LL-CP)	bar	3.97	0.91	2.32	2.29
Lower right earth chamber pressure (LR-CP)	bar	2.78	1.28	2.16	2.18
Center right earth chamber pressure (CR-CP)	bar	1.83	0.77	1.31	1.26
Screw conveyor pressure (SCP)	bar	80.09	24.76	47.04	46.55
Screw conveyor rotation speed (SCRS)	$r{\cdot}min^{-1}$	9.98	0.02	6.37	6.47
Screw conveyor torque (SCT)	kN·m	49.39	0.01	20.89	20.41
Advance speed (AS)	mm/min	75.89	5.37	51.69	53.23
Total thrust force (TTF)	kN	17877.24	10022.79	13695.38	13499.05
Penetration (P)	mm/r	61.89	6.13	38.03	38.82
Cutterhead torque (CT)	kN·m	4188.55	801.78	2307.38	2216.18
Cutterhead rotation speed (CRS)	$r{\cdot}min^{-1}$	1.82	0.88	1.37	1.32
Cutterhead pressure (CP)	bar	190.93	72.71	117.80	114.20

Note: The maximum, minimum, mean, and median values are the average of the four datasets.

of propelling jacks. Additionally, the earth chamber pressure is monitored across five divisions, as shown in Fig. 3b. The penetration (P) is defined as the ratio of the advancement distance to the corresponding propulsion time.

The EAMInfor model was trained on four datasets representing different geological conditions (see Table 1). This approach enables an analysis of the model's interpretability, allowing for an examination of whether the importance of various input features and specific points within input sequences differs across geological conditions. Geological parameters (e.g., cohesion, internal friction angle) were not included as model inputs, as these parameters remain constant within any given dataset. Including them could

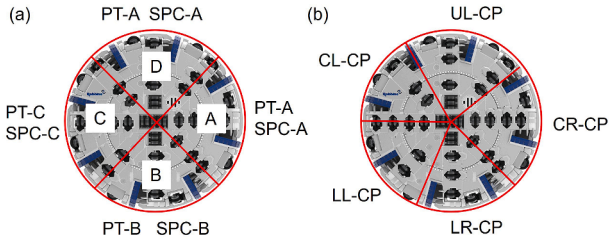

Fig. 3. Schematic diagram of shield tunneling parameters: (a) propulsion system and (b) earth chamber pressure.

lead the model to focus on irrelevant information, potentially wasting computational resources.

The prediction targets of the EAMInfor model are the positional deviations of the shield tunnel, including horizontal deviation of the shield head (HDSH), vertical deviation of the shield head (VDSH), horizontal deviation of the shield tail (HDST), and vertical deviation of the shield tail (VDST), as depicted schematically in Fig. 4. Significant horizontal deviations were observed at both the shield head and tail, measuring −157 mm and −164 mm, respectively, during shield tunneling in the MRS stratum, as shown in Fig. 5. In the FWG-SC stratum, the shield head experienced a notable vertical deviation of approximately −56 mm. The vertical deviation in the MRS stratum was greater than in the other strata.

4. Model development and performance

4.1. Model development

4.1.1. Evaluation metrics

The four datasets with different geological conditions presented in the previous section are divided into training, validation, and test sets in a ratio of 6:2:2. The EAMInfor model is trained and tested on each of the four datasets. The min-max scaling method was used to normalize the samples, scaling them to the range [0,1].

The model performance was assessed using mean absolute error (MAE), root mean squared error (RMSE), and coefficient of determination (R^2) parameters. The MAE provides the true value output in millimeters (mm) before the min-max scaling method is applied. Each experiment was repeated 10 times, and the average value was taken as the final performance metric. All experiments were conducted on a computing platform equipped with NVIDIA Tesla P100 and 64 GB of DDR4 RAM. The experimental environment utilized PyTorch 1.8 and Python 3.8.

4.1.2. Hyperparameter selection

The hyperparameter search space for the EAMInfor model is initially defined through pre-experiments, and the final hyperparameters are determined using the grid search method. Notably, SimAM is a parameter-free module. Since the four datasets, which represent varying geological conditions, are all derived from the same shield tunnel project, the optimal values of key hyperparameters remain consistent across the datasets, such as the reduction ratio, kernel size, number of heads in the multi-head attention mechanism, number of encoding layers, number of decoding layers, and learning rate. To ensure consistency in the model interpretability analysis, the optimal hyperparameters of the EAMInfor model are maintained uniformly across the four datasets, as detailed in Table 3.

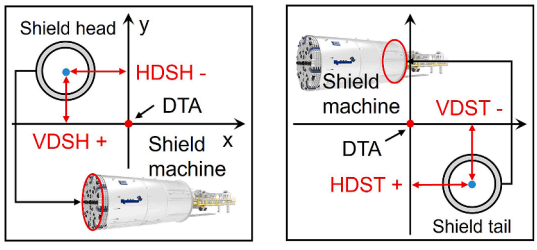

Fig. 4. Sketch of four parameters of the position deviation of the shield machine.

4.2. Performance of the EAMInfor model

Fig. 6 illustrates the performance of EAMInfor on the training and test sets across four distinct geological condition datasets. The close consistency between the model's performance on the training and test sets, along with its high prediction accuracy, indicates that the model avoids overfitting and demonstrates strong generalization capabilities.

4.2.1. Ablation experiment

Table 4 displays the results of the EAMInfor model on the test dataset. The results indicate that both the input and prediction sequence lengths significantly affect the model's performance. The EAMInfor model outperformed the Informer model across various input and prediction sequence lengths, highlighting the effectiveness of the channel, spatial, and SimAM attention mechanisms in enhancing model performance.

The EAMInfor model is optimal on the residual gravelly clay soil (RGC), fully weathered granite (FWG), and marine reclaimed sand (MRS) datasets with an input sequence length of 192 and a prediction sequence length of 64. Although the fully weathered granite mixed with silty clay (FWG-SC) dataset performed best with an input sequence length of 96, the performance difference between an input sequence length of 192 and the optimal length was negligible. In general, the channel attention mechanism of the EAMInfor model allows the model to focus on essential feature channels while filtering out irrelevant information, whereas the spatial attention mechanism enables it to concentrate on critical regions, thereby improving its ability to capture local information. Additionally, the SimAM module refines feature representation through similar computation, reducing noise and enhancing the model's generalization capabilities.

4.2.2. Comparison experiment

The comparison models include the PatchTST (Nie et al., 2022), FEDformer (Zhou et al., 2022), Reformer (Kitaev et al., 2020), Autoformer (Wu et al., 2021b), Transformer (Vaswani et al., 2017), GCN-LSTM (Zhang et al., 2024), and LSTM (Hochreiter and Schmidhuber, 1997) models. Table 5 shows the performance of the EAMInfor and the comparison models with input and prediction sequence lengths set to 192 and 64, respectively. The EAMInfor model significantly outperformed the other models, demonstrating its effectiveness and suitability for the prediction of shield tunneling position deviations.

The performance of PatchTST was slightly inferior to that of the EAMInfor model. This can be attributed to the lack of dedicated mechanisms in PatchTST for modeling channel and spatial information, which are likely critical for accurately predicting shield tunneling position deviations. Furthermore, across all four datasets, the GCN-LSTM model consistently outperformed the LSTM model, further highlighting the importance of extracting spatial features among channels in enhancing model performance.

Current machine learning models (Dai et al., 2023; Shen et al., 2022; Xu et al., 2023) for predicting shield position predominantly rely on recurrent neural network (RNN) architectures, such as LSTM and GRU models. However, comparative experimental

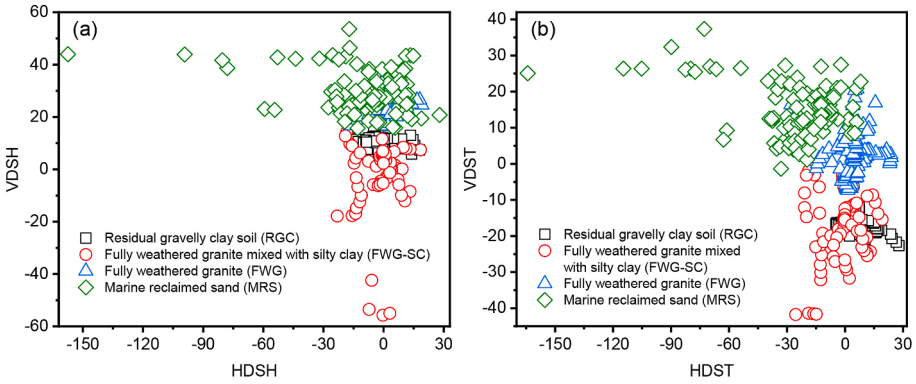

Fig. 5. Average positional deviation of concrete lining segment per ring under varying geological conditions: (a) HDSH and VDSH, and (b) HDST and VDST.

Table 3
Model hyperparameter search space and optimal combination.

Hyperparameter		Search space	Optimal hyperparameter
Hyperparameter of channel and spatial attention mechanism	Reduction ratio	[8, 16]	16
	Kernel size	[3, 7]	7
	Activation function	Sigmoid	Sigmoid
Hyperparameter of Informer	Number of heads in multi-head attention mechanism	[7, 8, 9]	8
	Number of encoding layers	[2, 3, 4]	3
	Number of decoding layers	[1, 2, 3]	2
	Dimension of model	[64, 128, 192]	128
	Activation function of distilling layer	ELU	ELU
Training parameter	Epoch	[100, 200, 300]	200
	Batch size	[48, 64, 96]	64
	Learning rate	[0.0001, 0.001, 0.01]	0.001
	Dropout	[0.05,0.1, 0.2]	0.05
	Loss function	MSE	MSE
	Optimizer	Adam	Adam

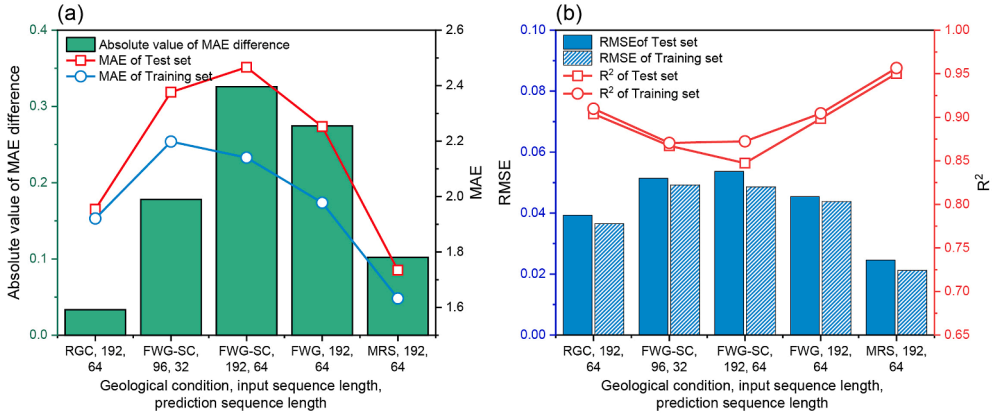

Fig. 6. Performance of EAMInfor on training and test set: (a) MAE, and (b) RMSE and R^2.

results indicate that models built on attention mechanisms and encoder-decoder architectures (e.g., EAMInfor, PatchTST, FEDformer, and Autoformer) generally outperform traditional models like LSTM in most scenarios. This superiority is attributed to the ability of Transformer-based models to effectively capture long-term dependencies in time-series data as the input and prediction sequence lengths increase. Moreover, the inherently sequential nature of RNN architecture limits their ability to leverage parallelized training, resulting in significantly increased computational complexity in long sequence time-series forecasting (LSTF) problems.

In the task of predicting shield tunneling position, increasing the input and prediction sequence lengths in machine learning models is crucial (Wang et al., 2023b). Extending the prediction sequence length allows the model to accurately forecast position deviations further into the future during the tunneling process. When the prediction sequence length exceeds the number of samples recorded in a single excavation step, the model can predict deviations for the next excavation step within the current step, providing more forward-looking guidance for construction. In addition, sufficiently long input sequences allow the model to capture more long-term dependency information from the data, which is particularly beneficial for identifying the dynamic patterns of low-dynamic features like SCRS and SCP.

Table 4
Model performance in different input and prediction sequence lengths.

Dataset	Input sequence length	Prediction sequence length	EAMInfor			Informer		
			MAE	RMSE	R^2	MAE	RMSE	R^2
RGC	96	32	1.9987	0.0432	0.8563	2.8312	0.1105	0.6852
	192	64	1.9543	0.0393	0.9039	2.3994	0.0613	0.7972
	256	128	2.1672	0.0476	0.8292	2.5532	0.0731	0.7622
FWG-SC	96	32	2.3763	0.0514	0.8673	2.9801	0.0861	0.6565
	192	64	2.4662	0.0537	0.8473	2.7945	0.0707	0.7399
	256	128	2.5892	0.0598	0.8177	3.1092	0.0922	0.6443
FWG	96	32	2.3982	0.0562	0.8542	2.5328	0.0849	0.7832
	192	64	2.2523	0.0454	0.8985	2.5921	0.0978	0.7632
	256	128	2.4562	0.0608	0.8076	2.8158	0.1398	0.7232
MRS	96	32	1.8761	0.0351	0.9242	2.3324	0.0557	0.8052
	192	64	1.7345	0.0246	0.9502	1.9674	0.0373	0.9152
	256	128	1.8460	0.0316	0.9308	2.4845	0.0705	0.7312

Table 5
Performance of EAMInfor and comparison models.

Model	Dataset	RGC	FWG-SC	FWG	MRS
EAMInfor	MAE	1.9543	2.4662	2.2523	1.7345
	RMSE	0.0393	0.0537	0.0454	0.0246
	R^2	0.9039	0.8473	0.8985	0.9502
PatchTST	MAE	2.1743	2.5082	2.4023	1.9787
	RMSE	0.0483	0.0543	0.0581	0.0377
	R^2	0.8302	0.8232	0.8312	0.9113
FEDformer	MAE	2.2334	2.6468	2.4958	2.4291
	RMSE	0.0591	0.0669	0.0816	0.0631
	R^2	0.8146	0.7766	0.7976	0.7827
Autoformer	MAE	2.7534	2.9532	2.8942	2.9503
	RMSE	0.1065	0.0764	0.1452	0.1222
	R^2	0.7071	0.6875	0.6941	0.6224
Transformer	MAE	2.4765	2.8897	2.5877	2.7798
	RMSE	0.0687	0.0726	0.0875	0.0867
	R^2	0.7805	0.7192	0.7756	0.6876
GCN-LSTM	MAE	2.8544	2.8977	2.6987	3.0882
	RMSE	0.1187	0.0745	0.0982	0.1309
	R^2	0.6575	0.6992	0.7601	0.6011
LSTM	MAE	2.8938	3.5741	3.9954	3.6579
	RMSE	0.1222	1.2759	0.1761	0.2198
	R^2	0.6282	0.6051	0.6388	0.4989

Note: Input sequence length = 192, prediction sequence length = 64.

5. Interpretability and limitations

5.1. Interpretability result of the DeepLIFT method

The contribution scores for input features and individual data points in the input sequence are calculated using the DeepLIFT method. Following the methodology described by Shrikumar et al. (2017), the reference activation for DeepLIFT was set to 0. A large absolute value of the contribution score calculated by DeepLIFT for a specific feature indicates its significant importance to the model's output.

5.1.1. Feature importance analysis

The absolute values of the contribution scores for the model's input features across different geological conditions are shown in Fig. 7. The colors in the figure denote different parameter types, and the numbers signify the importance of ranking each feature. We made the following observations based on the results in Fig. 7:

(1) The push thrust of push cylinder (PT-A, PT-B, PT-C, and PT-D) significantly affected the prediction results across different geological conditions, than other input features. The push thrust of push cylinder directly affects the machine's ability to overcome resistance from incoming strata while maintaining a stable tunneling speed and attitude. Specifically, the PT controls the shield machine's tunneling position deviations; for instance, operators can increase PT-A and decrease PT-C to rotate the shield machine to the left. Furthermore, excavating through different strata required appropriate adjustments to the thrust jack pressure based on the geological conditions.

Under all geological conditions, PT-B consistently exhibited higher values than PT-D, suggesting that the shield operator maintained a “head-up” position of the shield machine for an extended period to prevent the shield from plunging vertically, as illustrated in Fig. 8. The differences between PT-B and PT-D in the residual gravelly clay soil (RGC), fully weathered granite mixed with silty clay (FWG-SC), fully weathered granite (FWG), and marine reclaimed sand (MRS) strata are 54.5 kN, 108.8 kN, 91.6 kN, and 119.9 kN, respectively, indicating that the shield machine experienced the most pronounced “head-up” range in the MRS stratum. However, SPC, which also represents the state of the push cylinder, was found to be one of the least significant input features. Because the SPC increased from 600 mm to 1800 mm for each ring of the concrete lining (with one ring being 1200 mm wide), the variation in SPC was relatively stable and did not convey much information about changes in geological conditions or operator strategy.

(2) The importance ranking of AS and P was relatively consistent across different strata. In the RGC, FWG-SC, FWG, and MRS strata, AS ranked 9, 6, 8, and 7, respectively. There was a notable difference in the importance of P between the RGC and FWG-SC stratum. Additionally, the importance of AS and P in the same geological condition is usually opposite, i.e., one feature is high and the other is low. For instance, in the FWG-SC stratum, AS had the highest importance among the four strata, while P had the lowest. Conversely, in the RGC stratum, AS had higher importance than P. This pattern was due to the strong linear relationship between P and AS, as shown in Fig. 9. The Pearson correlation coefficient for AS and P in the RGC and FWG-SC strata were 0.9228 and 0.9731, respectively. These features may convey similar information; the model may focus on only one during model training to avoid overfitting and improve its generalization ability.

(3) The parameters related to earth chamber pressure are highly important, particularly those in the lower section (LR-CP, LL-CP). The lower earth chamber pressure is crucial for controlling the vertical attitude of the shield machine, whose primary function is to support the shield machine's weight and maintain its vertical position. The average of the

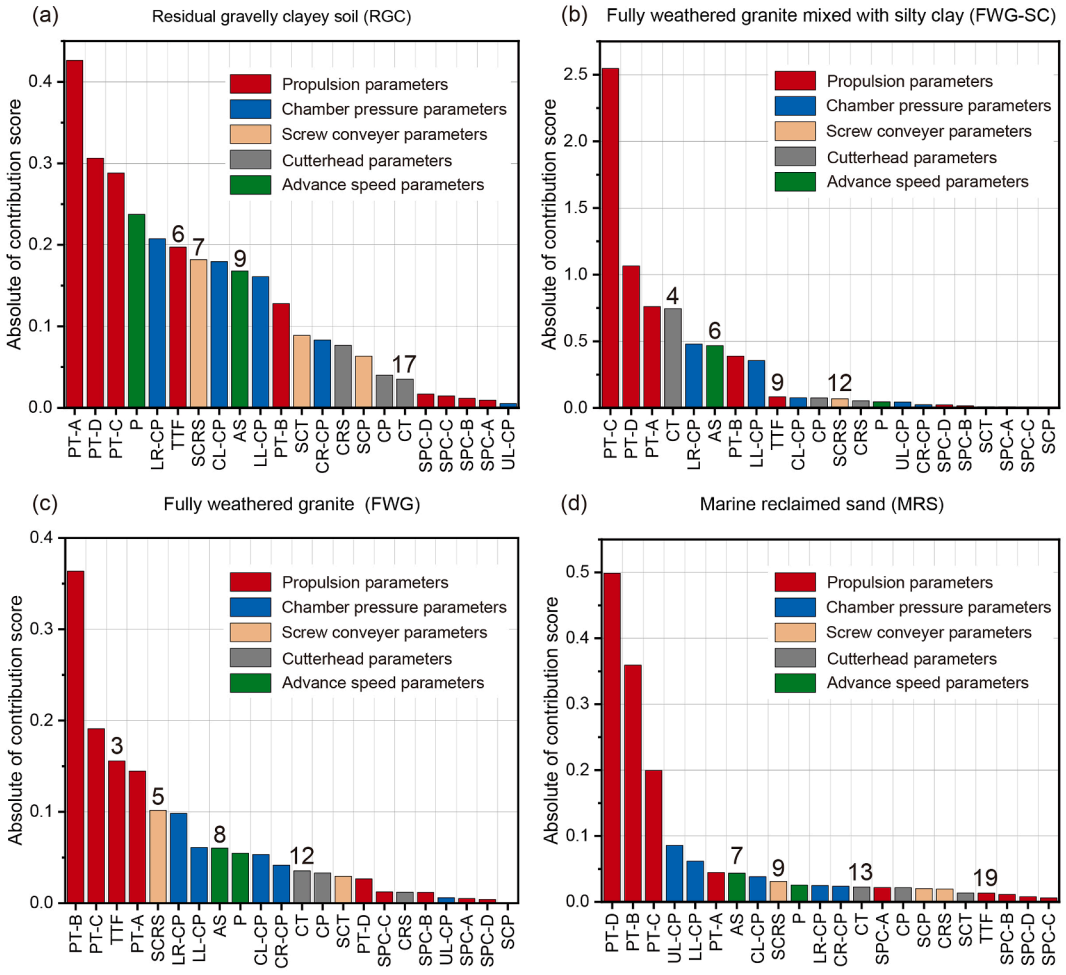

Fig. 7. Absolute value of feature contribution score in different geological conditions: (a) RGC, (b) FWG-SC, (c) FWG, and (d) MRS.

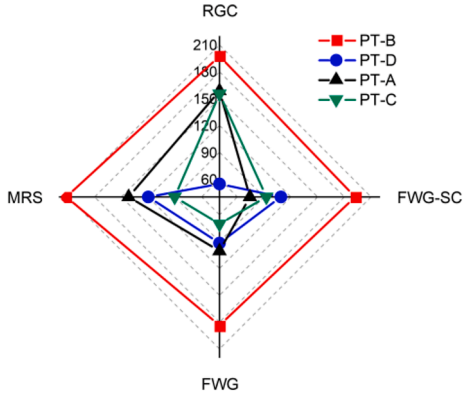

Fig. 8. Average of push thrust under different geological conditions.

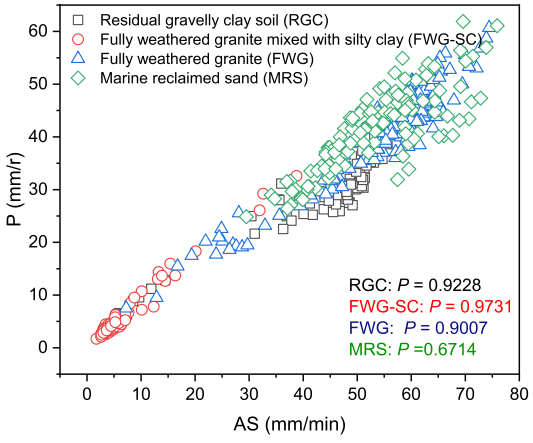

Fig. 9. Average AS and P of concrete lining segment per ring under varying geological conditions.

lower earth chamber pressures (LR-CP, LL-CP) in the four strata are higher than that in other parts, with LL-CP exceeding UL-CP, as shown in Fig. 10. Furthermore, the upper earth chamber pressure primarily prevents surface subsidence and collapse; its influence on the shield machine's position and attitude remains significant. In the marine reclaimed sand (MRS) dataset, UL-CP is the most important earth chamber pressure parameter. This could be due to the poor consolidation in the MRS strata and the significant surface subsidence at the site, which directly affects the vertical attitude of the shield machine. The large surface settlement leads to changes in upper earth chamber pressure, increasing the importance of UL-CP in the MRS stratum.

(4) Compared to other features, the importance of screw conveyor machine parameters (SCRS, SCP, and SCT) was

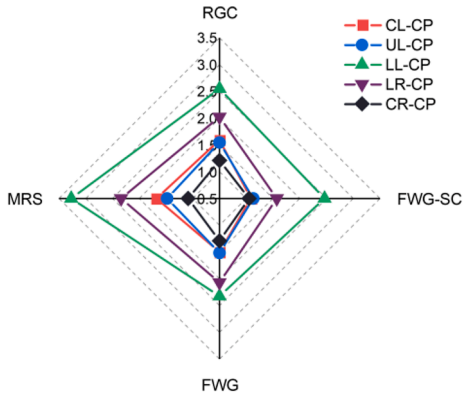

Fig. 10. Average of earth chamber pressure under different geological conditions.

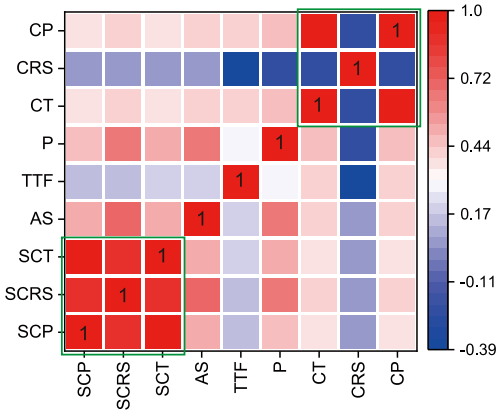

Fig. 11. Pearson correlation coefficients for some model input features.

relatively low. In addition, the importance of cutterhead parameters (CRS, CP, and CT) varied significantly across strata. Among the screw conveyor machine and cutterhead parameters categories, the model tends to focus on one key feature from each category. This is likely due to the high Pearson correlation coefficients among features within the same category, indicating that these features carry similar information, as shown in Fig. 11. For instance, SCRS consistently ranked higher in importance than SCP and SCT across all four geological conditions. In the FWG-SC, FWG, and MRS strata, CT ranked higher than CRS and CP. In summary, the model tends to focus on only one of several features with strong linear correlations, as these features often convey the same information.

5.1.2. Importance of data points in the input sequence

Figs. 12 and 13 illustrate the DeepLIFT contribution scores for different sample points within the input sequence across datasets under varying geological conditions. The x-axis represents the indices of different points in the input sequence (with an input sequence length of 192), while the y-axis depicts the absolute values of the contribution scores for each sample point. The results indicate that the importance of data points at different positions in the input sequence varied significantly among the four strata. In the RGC and FWG strata, the contribution scores showed distinct trends: scores from data points 1–52 increased, those from data points 53–108 fluctuated, scores from data points 108–182 increased and peaked around 182, and scores from data points

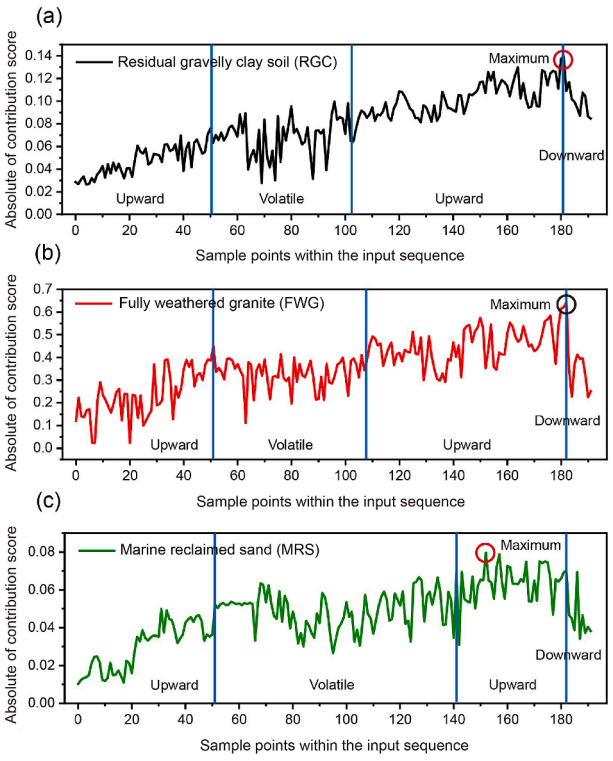

Fig. 12. Absolute contribution score of input sequence data point in different geological conditions: (a) RGC, (b) FWG, and (c) MRS.

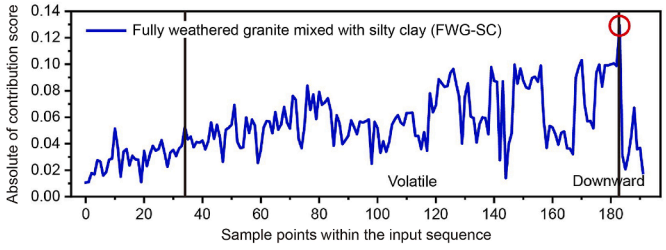

Fig. 13. Absolute contribution score of input sequence data point in FWG-SC stratum.

183-192 decreased (Fig. 12a and b). The RGC and FWG strata achieved their maximum contribution scores in the 181th and 182nd data points, respectively.

The fluctuation phase was more pronounced in the MRS stratum compared to the RGC and FWG strata, spanning from data points 53 to 142 (Fig. 12c). The contribution patterns of data points in the input sequence across all three strata (RGC, FWG, and MRS) exhibited a consistent four-phase trajectory: an initial upward trend, followed by a period of volatility, then another upward phase, before finally trending downward. This declining contribution of data points toward the end of the input sequence explains why the model achieved superior performance at an input sequence length of 192 compared to 256 (see Table 4).

In the FWG-SC stratum, the contribution score trend differed from the other strata. There was an upward trend in contribution scores around data points 1–38, followed by a fluctuating trend around data points 50–97. However, from data points 97–122, the contribution scores did not rise again but instead fluctuated greatly. At around data point 182, the contribution scores began to decline (Fig. 13). This pattern explains why the optimal input sequence length for the EAMInfor model in the FWG-SC stratum

dataset was 96, rather than 192 (see Table 4). This is because the importance of the data points did not rise again after the 97th point, as in other strata.

The contribution scores trend for each data point in the input sequence, as derived using the DeepLIFT method, aligned with the changes observed in model performance. This consistency reinforces the conclusions drawn from the DeepLIFT method. Due to the significant variation in the importance of data points at different positions within the input sequence length, selecting the appropriate input sequence length is crucial for achieving optimal performance in time series forecasting models.

5.2. Validation of interpretability results

Since the DeepLIFT method has rarely been applied to interpretability analysis in ML prediction models for shield tunneling position deviation. To validate its interpretability results, we employ a widely adopted test using various input parameter combinations alongside the SHAP method. Notably, the SHAP method does not rely on the EAMInfor model for predictions but instead builds an independent ML prediction model. This approach ensures that the interpretability results remain unaffected by the specific characteristics of the EAMInfor model. Furthermore, it also allows for an evaluation of the generalizability of DeepLIFT's interpretability results, ensuring their validity beyond the constraints of any single model.

5.2.1. Tests with different input feature combinations

To assess the performance of the DeepLIFT method for ranking feature importance, we conducted tests using various input feature combinations using the best-performing marine reclaimed sand (MRS) dataset (see Table 6). It is worth noting that for the EAMInfor model with all 22 input features in the MRS dataset, the MAE, RMSE, and R^2 were 1.7345, 0.0246, and 0.9502, respectively (see Table 4). The stronger performance of the model with different combinations of input features suggests a higher significance of the features within those specific input feature combinations.

The results for Groups 1 to 2 indicate that incorporating the input features SPC-A, SPC-B, SPC-C, and SPC-D led to a decline in model performance. This phenomenon could be attributed to a high correlation between newly added features and existing ones, which may result in a multicollinearity issue. In this case, even if the new features contain useful information, they may lead to model instability due to redundancy with other features. Furthermore, if there is a high linear correlation between the newly added features, it can result in the model assigning too much weight to these features while ignoring the more important ones. As shown in Fig. 14, the Pearson correlation coefficients between SPC-A, SPC-B, SPC-C, and SPC-D are close to 1, indicating a very strong linear correlation. Furthermore, the linear correlation between SPC and PT is significantly larger than the linear correlation between SPC and the five earth chamber pressure features. This explains why Group 2 shows a significant decrease in performance compared to Group 1. Groups 3–5 were modeled by adding “SCT, SCRS, SCP”, “CT, CRS, CP”, and “AS, TTF, P″ to Group 1, respectively. Adding “CT, CRS, CP” and “AS, TTF, P″ improved the model performance. However, adding “SCT, SCRS, SCP” resulted in a decrease in model performance.

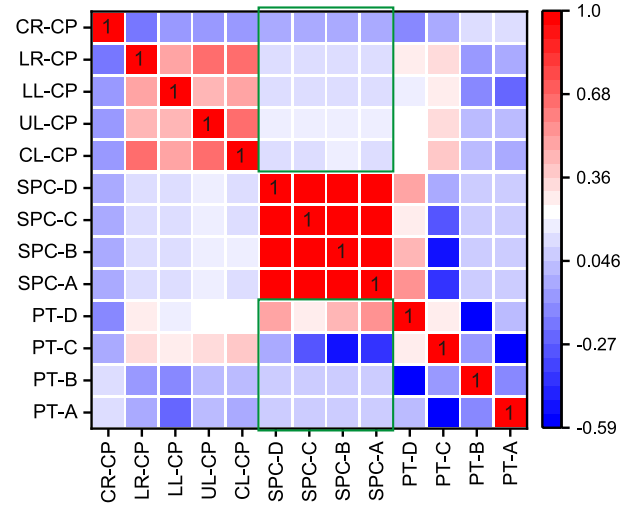

Fig. 14. Average of Pearson's correlation coefficient in model input features.

5.2.2. Feature importance analysis based on the SHAP method

The prediction accuracy of the SHAP method predictor significantly affects the confidence of its feature interpretation. ML models such as extreme gradient boosting (XGBoost), light gradient boosting machine (LightGBM), categorical boosting (CatBoost), random forest (RF), adaptive boosting (AdaBoost), and gradient boosting decision tree (GBDT) are selected as candidate models for the SHAP method predictors. The dataset division, hyperparameter selection, and evaluation metrics are consistent with Section 4.1.

The CatBoost model outperforms the other models on the residual gravelly clay soil (RGC), fully weathered granite (FWG), and marine reclaimed sand (MRS) datasets, as shown in Table 7. The XGBoost model outperforms the other models on the fully weathered granite mixed with silty clay (FWG-SC) dataset. In each dataset, the best-performing model was selected as the predictor for the SHAP method. The hyperparameters for the CatBoost and XGBoost models are detailed in Table 8.

The Shapley value of the input features calculated by the SHAP method is shown in Fig. 15. The numbers in the figure represent the importance rankings derived from the DeepLIFT method. The results of the SHAP method were broadly like those of the DeepLIFT method. In all four geological conditions, push thrust of push cylinder (PT-A, PT-B, PT-C, and PT-D) showed strong importance, while the stroke length of the push cylinder (SPC-A, SPC-B, SPC-C, and SPC-D) ranked among the least important features. The earth chamber pressure parameters (CL-CP, UL-CP, LL-CP, LR-CP, and CR-

Table 6
Performance of different input parameter combinations in the test set.

Group number	Feature name	MAE	RMSE	R^2
1	PT-A, PT-B, PT-C, PT-D, CL-CP, UL-CP, LL-CP, LR-CP, CR-CP	2.5321	0.0519	0.8714
2	PT-A, PT-B, PT-C, PT-D, CL-CP, UL-CP, LL-CP, LR-CP, CR-CP, SPC-A, SPC-B, SPC-C, SPC-D	2.6960	0.0569	0.8281
3	PT-A, PT-B, PT-C, PT-D, CL-CP, UL-CP, LL-CP, LR-CP, CR-CP, SCT, SCRS, SCP	2.6299	0.0527	0.8435
4	PT-A, PT-B, PT-C, PT-D, CL-CP, UL-CP, LL-CP, LR-CP, CR-CP, CT, CRS, CP	2.3782	0.0494	0.8892
5	PT-A, PT-B, PT-C, PT-D, CL-CP, UL-CP, LL-CP, LR-CP, CR-CP, AS, TTF, P	2.2882	0.0472	0.8952

Note: Prediction sequence length = 64, Input sequence length = 192.

Table 7
Performance of different ML models in different geological conditions datasets.

Model	RGC dataset			FWG-SC dataset			FWG dataset			MRS dataset		
	MAE	RMSE	R^2	MAE	RMSE	R^2	MAE	RMSE	R^2	MAE	RMSE	R^2
XGBoost	2.9902	0.1241	0.6046	3.0021	0.0896	0.6513	3.0421	0.1471	0.6815	2.9982	0.1253	0.6177
LightGBM	3.2867	0.1365	0.5812	3.6677	1.3198	0.5962	3.7721	0.1706	0.6492	3.4821	0.1571	0.5682
CatBoost	2.7621	0.1092	0.7042	3.1118	0.0944	0.6416	2.7981	0.0992	0.7597	2.6555	0.0819	0.7002
RF	3.6582	0.1427	0.5762	4.0721	1.3917	0.5402	4.7621	0.1992	0.5965	3.3562	0.1456	0.5921
AdaBoost	2.8971	0.1233	0.6194	3.3452	0.1025	0.6378	2.8821	0.1422	0.7093	2.8976	0.0931	0.6592
GBDT	3.8762	0.1552	0.5582	3.7721	1.3476	0.5882	3.2621	0.1598	0.6708	3.5421	0.1716	0.5377

Table 8
Hyperparameters of the CatBoost and XGBoost models.

Model	Hyperparameters	Search space	Optimal hyperparameters
CatBoost	Max depth	[6, 9, 12]	9
	Iterations	[600, 900, 1200]	900
	Regularization	[1, 3, 5]	3
	Learning rate	[0.03, 0.02, 0.01]	0.02
XGBoost	Max depth	[6, 9, 12]	9
	Subsample	[0.8, 0.9, 1]	0.9
	Number of estimators	[600, 900, 1200]	1200
	Minimum splitting loss	[0.3, 0.5]	0.3
	Regularization	[1, 3, 5]	3
	Learning rate	[0.1, 0.05, 0.01]	0.05

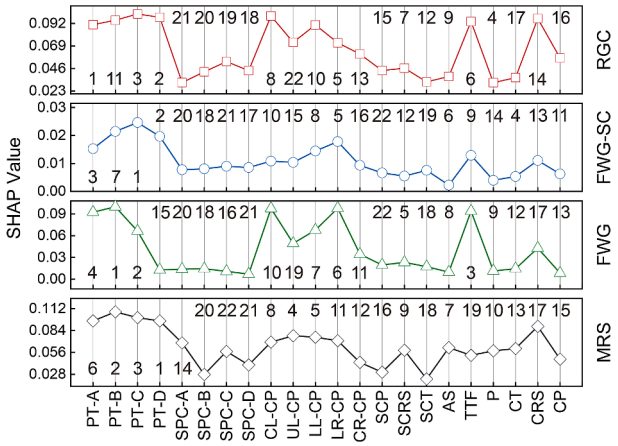

Fig. 15. Feature important results of the SHAP method. The numbers in the figure represent the feature of importance rankings derived from the DeepLIFT method.

CP) also exhibited high feature importance. The importance of the lower earth chamber pressure was greater than that of the earth chamber pressure at other locations.

Across the four strata, SCRS, AS, and P generally showed lower SHAP importance rankings than their DeepLIFT rankings, while CRS consistently ranked higher with SHAP. This discrepancy may be attributed to the differences in algorithmic mechanisms or correlations within the temporal data. Since the SHAP predictor does not account for temporal correlation within the datasets, it overlooks potential relationships between input features and shield position parameters over time. This limitation can result in rankings of certain features being of reduced importance. For example, SCRS indirectly affects earth chamber pressure by controlling the soil content within the cutterhead's soil compartment. As the change in earth chamber pressure resulting from alterations in SCRS is delayed, the SHAP-based predictor struggles to effectively capture this temporal correlation, leading to a lower SHAP feature importance for SCRS compared to the DeepLIFT method.

In contrast, CRS has a more direct impact on shield position deviation. For example, changes in CRS immediately affect cutting efficiency and tunneling stability. Due to its inability to account for temporal dependencies, CatBoost tends to emphasize such instantaneous effects, leading SHAP to assign a higher importance score to CRS. In summary, the DeepLIFT and SHAP analyses of input feature importance produced largely similar feature importance rankings. This highlights the applicability and reliability of the DeepLIFT method for deep learning-based predictive modeling of shield tunneling position deviations.

5.3. Limitations

Although the EAMInfor model achieves high prediction accuracy, the model has numerous parameters when handling long sequence data. The high number of parameters in the EAMInfor model may compromise its ability to generalize effectively to other shield tunneling projects. Furthermore, the EAMInfor model does not include geological condition information as an input feature. This is because the objective of this study is to examine whether the focus of the shield tunneling position deviation prediction model varies under different geological conditions. Therefore, all four datasets used in this study are derived from a single homogeneous stratum, where the geological parameters within each dataset remain constant.

The reference activation in the DeepLIFT method is a critical factor affecting the outcomes. Determining the optimal reference activation can be challenging in practical applications. Moreover, the DeepLIFT interpretation is heavily dependent on the model's internal structure, particularly the activation functions and interlayer connections. Consequently, different model structures may require specific adjustments and optimizations for DeepLIFT. While both DeepLIFT and SHAP methods produce similar conclusions regarding feature importance, discrepancies exist in their rankings. DeepLIFT's analysis is often preferred due to the superior performance of the EAMInfor model compared to CatBoost and XGBoost. However, model performance alone does not guarantee the credibility of interpretability results. A deeper understanding of how input features influence shield position deviations remains

lacking, highlighting the current limitations of interpretability in this context.

6. Conclusions

A novel explainable deep learning framework has been proposed, which incorporates the Informer model enhanced with channel, spatial, and SimAM attention mechanisms (EAMInfor) and deep learning important features (DeepLIFT). The framework was applied to datasets encompassing four different geological conditions from the Xiamen metro line 3 project in China. Furthermore, the interpretability of the model is further supported by various experiments utilizing different combinations of input features and Shapley additive explanations (SHAP) methods. The main conclusions are as follows:

(1) The performance of the Informer model is significantly improved by adding channels, spatial, and SimAM attention mechanisms. The optimal input and prediction sequence lengths for the EAMInfor model were 192 and 64, respectively. Furthermore, the EAMInfor model outperformed comparison models such as PatchTST, FEDformer, Autoformer, and LSTM.
(2) The push thrust of push cylinder and the earth chamber pressure significantly impacted the prediction of shield position deviation in different geologic conditions datasets, whereas the stroke length of the push cylinder exhibited lower importance. Lower earth chamber pressure was more important than earth chamber pressure at other locations in RGC, FWG-SC, and FWG datasets, while upper left earth chamber pressure is the most important earth chamber pressure parameter in the MRS dataset.
(3) The importance of data points at different locations within the input sequence can vary significantly. Furthermore, the significance of data points within the input sequence exhibits distinct patterns: in homogeneous strata, it tends to increase, fluctuate, increase again, and then decrease, whereas in composite strata, it displays continuous fluctuation throughout. Selecting an appropriate input sequence length can further enhance model performance.
(4) The feature importance analysis results obtained using the SHAP method and tests with different input feature combinations were generally consistent with those of the DeepLIFT method. However, because SHAP's predictor does not incorporate temporal dependencies in time series forecasting, features such as screw conveyor rotation speed and advance speed were assigned lower importance compared to the DeepLIFT method. Overall, the SHAP results further reinforce the credibility and reliability of the DeepLIFT method.

In summary, as the interpretability of deep learning models for predicting shield tunnel position deviation increases, tunnel construction managers can rely on the model decisions for more reliable and effective shield position deviation predictions. Our future work will focus on intrinsic methods within XAI approaches to explore the underlying reasons and mechanisms behind the model's predictions through its internal structure. In addition, integrating the physical mechanics of shield position deviation into machine learning may help further unlock the “black box” of these models. To improve engineering applicability and computational efficiency, techniques such as knowledge distillation and transfer learning will also be explored to streamline model parameters.

CRediT authorship contribution statement

Jiajie Zhen: Writing – review & editing, Writing – original draft, Software, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Fengwen Lai:** Writing – review & editing, Project administration, Methodology, Funding acquisition. **Ming Huang:** Writing – review & editing, Validation, Project administration, Funding acquisition, Data curation. **Junjie Zheng:** Writing – review & editing, Validation, Supervision. **Jim S. Shiau:** Writing – review & editing. **Ping Wang:** Project administration, Data curation. **Jinhuo Zheng:** Project administration, Supervision, Validation, Writing – review & editing.

Declaration of competing interest

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

Acknowledgment

This study was supported by the National Natural Science Foundation of China (Grant Nos. 52378392, 52408356) and the Foal Eagle Program Youth Top-notch Talent Project of Fujian Province, China (Grant No. 00387088).

References

Cao, Y., Luo, W., Xue, Y., Lin, W., Zhang, F., 2024. Model-based offline reinforcement learning framework for optimizing tunnel boring machine operation. Undergr. Space 19, 47–71.
Chen, C., Fan, L., 2023. An attribution deep learning interpretation model for landslide susceptibility mapping in the three gorges reservoir area. IEEE Trans. Geosci. Rem. Sens. 61, 3000515.
Chen, H., Gomez, C., Huang, C.M., Unberath, M., 2022. Explainable medical imaging AI needs human-centered design: guidelines and evidence from a systematic review. npj Digit. Med. 5, 156.
Chen, L., Tian, Z., Zhou, S., Gong, Q., Di, H., 2024. Attitude deviation prediction of shield tunneling machine using time-aware LSTM networks. Transp. Geotech. 45, 101195.
Chen, W., Tan, X.Y., Yang, J., 2025. Review of state-of-the-art in structural health monitoring of tunnel engineering. Smart Undergr. Eng.
Dai, Z.Y., Li, P.A., Zhu, M.Q., Zhu, H.H., Liu, J., Zhai, Y.X., Fan, J., 2023. Dynamic prediction for attitude and position of shield machine in tunneling: a hybrid deep learning method considering dual attention. Adv. Eng. Inform. 57, 102032.
Duan, S., Song, Z., Shen, J., Xiong, J., 2024a. Prediction for underground seismic intensity measures using conditional generative adversarial networks. Soil Dynam. Earthq. Eng. 180, 108619.
Duan, S., Zhao, G., Jiang, Q., Xiong, J., Sun, Y., Kou, Y., Qiu, S., 2024b. Multi-index fusion database and intelligent evaluation modelling for geostress classification. Tunn. Undergr. Space Technol. 149, 105802.
Fang, Y.R., Li, X.G., Liu, H.Z., Hao, S.N., Yi, Y., Guo, Y.D., Li, H.Y., 2023. Intelligent real-time identification technology of stratum characteristics during slurry TBM tunneling. Tunn. Undergr. Space Technol. 139, 105216.
Fu, X., Wu, M., Ponnarasu, S., Zhang, L., 2023. A hybrid deep learning approach for dynamic attitude and position prediction in tunnel construction considering spatio-temporal patterns. Expert Syst. Appl. 212, 118721.
He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. Proc. IEEE Conf. Comput. Vis. Pattern Recognit 770–778.
Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. Neural Comput. 9 (8), 1735–1780.
Hu, M., Zhang, H., Wu, B., Li, G., Zhou, L., 2022. Interpretable predictive model for shield attitude control performance based on XGboost and SHAP. Sci. Rep. 12 (1), 18226.
Huang, M., Lu, Y., Zhen, J., Lan, X., Xu, C., Yu, W., 2023. Analysis of face stability at the launch stage of shield or TBM tunnelling using a concrete box in complex urban environments. Tunn. Undergr. Space Technol. 135, 105067.
Kitaev, N., Kaiser, Ł., Levskaya, A., 2020. Reformer: the efficient transformer. Int. Conf. Mach. Learn. (PMLR).
Lai, F., Liu, S., Shiau, J., Liu, M., Cai, G., Huang, M., 2025. Data-driven modeling for evaluating deformation of a deep excavation near existing tunnels. Undergr. Space.
Lai, F., Shiau, J., Keawsawasvong, S., Chen, F., Banyong, R., Seehavong, S., 2023. Physics-based and data-driven modeling for stability evaluation of buried structures in natural clays. J. Rock Mech. Geotech. Eng. 15 (5), 1248–1262.
Lai, F., Tschuchnigg, F., Schweiger, H.F., Liu, S., Shiau, J., Cai, G., 2024. A numerical

study of deep excavations adjacent to existing tunnels: integrating CPTU and SDMT to calibrate soil constitutive model. Can. Geotech. J. 62, 1–23.
Lin, S., Dong, M., Cao, X., Liang, Z., Guo, H., Zheng, H., 2024. The pre-trained explainable deep learning model with stacked denoising autoencoders for slope stability analysis. Eng. Anal. Bound. Elem. 163, 406–425.
Linardatos, P., Papastefanopoulos, V., Kotsiantis, S., 2021. Explainable AI: a review of machine learning interpretability methods. Entropy 23, 1.
Liu, Z., Li, L., Fang, X., Qi, W., Shen, J., Zhou, H., Zhang, Y., 2021. Hard rock tunnel lithology prediction with TBM construction big data using a global attention mechanism based LSTM network. Autom. ConStruct. 125, 103647.
Lu, D., Liu, Y., Kong, F., He, X., Zhou, A., Du, X., 2024. A novel Bi-LSTM method fusing current and historical data for tunnelling parameters of shield tunnel. Transp. Geotech. 49, 101402.
Lundberg, S.M., Lee, S.I., 2017. A unified approach to interpreting model predictions. Int. Conf. Mach. Learn. (PMLR). 30.
Nie, Y., Nguyen, N.H., Sinthong, P., Kalagnanam, J., 2022. A time series is worth 64 words: long-Term forecasting with transformers. 11th Int. Conf. Learn. Represent.
Ortigossa, E.S., Gonçalves, T., Nonato, L.G., 2024. EXplainable artificial intelligence (XAI)–from theory to methods and applications. IEEE Access 12, 80799–80846.
Schubert, E., Sander, J., Ester, M., Kriegel, H.P., Xu, X., 2017. DBSCAN revisited, revisited: why and how you should (still) use DBSCAN. ACM Trans. Database Syst. 42, 1–21.
Seu, K., Kang, M.-S., Lee, H., 2022. An intelligent missing data imputation techniques: a review. JOIV: International Journal on Informatics Visualization 6, 278–283.
Shen, S.L., Elbaz, K., Shaban, W.M., Zhou, A., 2022. Real-time prediction of shield moving trajectory during tunnelling. Acta Geotech 17, 1533–1549.
Shen, X., Chen, X., Bao, X., Zhou, R., Zhang, G., 2023. Real-time prediction of attitude and moving trajectory in shield tunneling based optimal input parameter combination using random forest deep learning method. Acta Geotech 18, 1–21.
Shrikumar, A., Greenside, P., Kundaje, A., 2017. Learning important features through propagating activation differences. Int. Conf. Mach. Learn (PMLR). 70, 3145–3153.
Tan, X.Y., Chen, W., Fan, L., Ye, J., Du, B., 2025a. Spatial dynamic early warning of different positions in underwater tunnel driven by real-time monitoring data. Struct. Control Health Monit. 2025, 5397749.
Tan, X., Chen, W., Tan, X., Fan, C., Mao, Y., Cheng, K., Du, B., 2025b. Missing data imputation in tunnel monitoring with a spatio-temporal correlation fused machine learning model. J. Civ. Struct. Health Monit. 15, 1337–1348.
Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I., 2017. Attention is all you need. Adv. Neural Inf. Process. Syst. 30.
Wang, G., Shan, Y., Detmann, B., Lin, W., 2024. Physics-informed Neural Network (PINN) model for predicting subgrade settlement induced by shield tunnelling beneath an existing railway subgrade. Transp. Geotech. 49, 101409.
Wang, K., Wu, X., Zhang, L., Song, X., 2023a. Data-driven multi step robust prediction of TBM attitude using a hybrid deep learning approach. Adv. Eng. Inform. 55, 101854.
Wang, P., Kong, X., Guo, Z., Hu, L., 2019. Prediction of axis attitude deviation and deviation correction method based on data driven during shield tunneling. IEEE Access 7, 163487–163501.
Wang, Y., Zhao, J., Jiang, K., Zhou, Q., Kang, Z., Chen, C., Zhang, H., 2023b. Prediction of TBM operation parameters using machine learning models based on BPSO. Adv. Eng. Inform. 56, 101955.
Wu, H., Xu, J., Wang, J., Long, M., 2021a. Autoformer: decomposition transformers with auto-correlation for long-term series forecasting. Adv. Neural Inf. Process. Syst. 34, 22419–22430.
Wu, Z., Wei, R., Chu, Z., Liu, Q., 2021b. Real-time rock mass condition prediction with TBM tunneling big data using a novel rock-machine mutual feedback perception method. J. Rock Mech. Geotech. Eng. 13 (16), 1311–1325.
Xiao, H., Chen, Z., Cao, R., Cao, Y., Zhao, L., Zhao, Y., 2022. Prediction of shield machine posture using the GRU algorithm with adaptive boosting: a case study of chengdu subway project. Transp. Geotech. 37, 100837.
Xiao, H., Xing, B., Wang, Y., Yu, P., Liu, L., Cao, R., 2021. Prediction of shield machine attitude based on various artificial intelligence technologies. Appl. Sci. 11 (21), 10264.
Xu, J., Bu, J., Qin, N., Huang, D., 2024. SCA-MADRL: multiagent deep reinforcement learning framework based on state classification and assignment for intelligent shield attitude control. Expert Syst. Appl. 235, 121258.
Xu, J., Zhang, Z., Zhang, L., Liu, D., 2023. Predicting shield position deviation based on double-path hybrid deep neural networks. Autom. ConStruct. 148, 104775.
Yang, L., Zhang, R.Y., Li, L., Xie, X., 2021. Simam: a simple, parameter-free attention module for convolutional neural networks. Int Conf Mac. Learn (PMLR). 11863–11874.
Zhang, D., Shen, Y., Huang, Z., Xie, X., 2022. Auto machine learning-based modelling and prediction of excavation-induced tunnel displacement. J. Rock Mech. Geotech. Eng. 14 (4), 1100–1114.
Zhang, L., Guo, J., Fu, X., Tiong, R.L.K., Zhang, P., 2024. Digital twin enabled real-time advanced control of TBM operation using deep learning methods. Autom. ConStruct. 158, 105240.
Zhang, P., Chen, R.P., Wu, H.N., 2019. Real-time analysis and regulation of EPB shield steering using random forest. Autom. ConStruct. 106, 102860.
Zheng, Z., Luo, K., Tan, X., Jia, L., Xie, M., Xie, H., Jiang, L., Gong, G., Yang, H., Han, D., 2024. Autonomous steering control for tunnel boring machines. Autom. ConStruct. 159, 105259.
Zhou, C., Gao, Y., Chen, E.J., Ding, L., Qin, W., 2023. Deep learning technologies for shield tunneling: challenges and opportunities. Autom. ConStruct. 154, 104982.
Zhou, C., Xu, H., Ding, L., Wei, L., Zhou, Y., 2019. Dynamic prediction for attitude and position in shield tunneling: a deep learning method. Autom. ConStruct. 105, 102840.
Zhou, T., Ma, Z., Wen, Q., Wang, X., Sun, L., Jin, R., 2022. Fedformer: frequency enhanced decomposed transformer for long-term series forecasting. Int. Conf. Mach. Learn (PMLR). 27268–27286.

Dr Ming Huang is currently Professor of Fuzhou University, China. He is Vice Dean of the Civil Engineering College (responsible for overall administrative work). He is a member of the Chinese Society for Rock Mechanics and Engineering. His research interests include (1) multiscale mechanical properties of bio-cemented soils and EICP/MICP-based environment-friendly ground reinforcement methods, (2) soil conditioning and muck recycling of shield tunneling, and (3) mechanical effect analysis and disaster prevention of tunnel construction in complex environments.