



Prediction of surface settlement during shield tunnel boring based on variational mode decomposition and machine learning models

Na He, Pengfei Li , Mingju Zhang, Chenhe Ge ^{*}

Key Laboratory of Urban Security and Disaster Engineering, Ministry of Education, Beijing University of Technology, Beijing 100124, China

ARTICLE INFO

Keywords:

Settlement prediction
Variational mode decomposition
Bidirectional long and short-term memory
neural networks
Sparrow search algorithm
Attention mechanism

ABSTRACT

Surface settlement caused by shield tunnel boring can damage existing structures and the surrounding environment. Machine learning is currently playing a major role in predicting surface settlement. However, most of the models only consider hyperparameter optimization and input parameter weight assignment, but fail to adequately capture the local abrupt change features of surface settlement time series data. This paper proposes a VMD-SSA-ATT-BiLSTM model (hereinafter referred to as VSA-BiLSTM) for predicting surface settlement. The proposed model combines VMD (Variational Mode Decomposition), SSA (Sparrow Search Algorithm), ATT (Attention Mechanism), and BiLSTM (Bidirectional Long Short-Term Memory). First, the surface settlement data was decomposed into multiple simple subsequences using VMD; Secondly, the SSA is introduced to optimize the hyper-parameters of the BiLSTM; ATT is then introduced to assign the feature weights extracted by BiLSTM; Finally, the SSA-ATT-BiLSTM model was used to independently train and predict the subsequences of each VMD decomposition, and the predictions of all subsequences were subsequently summarized to produce the final surface settlement prediction. The model was validated using the case of the Shenzhen Dayawan Intercity Tunnel, and compared with four mainstream prediction models. The Taylor diagram and multiple evaluation indicators were used to analyze the prediction performance of the model, and the SHAP (SHapley Additive explanations) visualization method was applied to study the contribution of all input features to the model's prediction results. The results show that: Compared with existing models, this model not only effectively achieves data denoising but also exhibits superior predictive performance and high robustness in surface settlement prediction, providing a reference for early warning of surface settlement disasters during construction.

1. Introduction

Shield tunneling inevitably disturbs surrounding soils; this disturbance induces stress redistribution within the soil mass, ultimately resulting in ground settlement and deformation (Song et al., 2025). When the surface settlement caused by shield tunnel boring exceeds the permissible range, it will cause damage to the surrounding existing structures and even lead to various types of engineering disasters, seriously jeopardizing the lives and properties of residents. Accurate settlement predictions can help mitigate structural damage, reduce costs and improve the safety of urban tunneling projects. Therefore, predicting the surface settlement induced by shield tunneling is of significant engineering importance.

Currently, the main methods for predicting surface settlement triggered by shield tunneling are: theoretical analysis, numerical simulation, and machine learning. Theoretical analysis methods use

mathematical models to predict surface settlement, which are computationally efficient but rely on assumptions, have limited accuracy, and are only applicable to simple working conditions. Numerical simulation (such as FLAC and ABAQUS) has a wide range of applications and provides intuitive results, but the modeling process is complex and relies on a large number of parameters to ensure accuracy. Machine learning is a data-driven approach to build mathematical models based on historical measured data for surface settlement prediction, its core is to learn the regularities of data through algorithms, it includes methods such as regression analysis (Tian et al., 2021), random forests (Wang et al., 2023), support vector machines (Xu et al., 2023), and neural networks (Wang et al., 2023; Gan, 2025), among other methods. With the continuous development of computer technology, neural networks are gradually being applied to surface settlement prediction tasks (Ju et al., 2024). Neural networks possess powerful nonlinear fitting capabilities, and with reasonable model structures, they can achieve accurate surface

^{*} Corresponding address.

E-mail address: 2996845330@qq.com (C. Ge).

<https://doi.org/10.1016/j.tust.2025.107216>

Received 29 April 2025; Received in revised form 4 October 2025; Accepted 23 October 2025

Available online 31 October 2025

0886-7798/© 2025 Elsevier Ltd. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

settlement predictions without relying on complex mathematical or physical models (Liu et al., 2024; Nie et al., 2025). Table 1 shows the different machine learning models used for prediction in recent years, along with their advantages and disadvantages.

The surface settlement data show non-stationarity and short-term abruptness in the time series, it also contains a lot of noise, which will produce a large error if the raw settlement data are used directly for prediction (Duan, 2025; Wang et al., 2020). Current conventional surface settlement prediction models, due to their structural limitations, cannot identify and eliminate noise, so it is not possible to avoid the impact of extreme noise on the model prediction performance, which leads to a decrease in the accuracy of the model prediction and increases the engineering risk. (Zhang et al., 2024; Wang et al., 2025) found that when collecting the construction parameters of shield tunnel, the data would be noisy due to the influence of the complex environment and the factor of manual error, so the VMD was introduced into the hybrid model 1DCNN-LSTM-ATT, which reduced the noise of the data before the model predicted the parameters of the shield tunnel construction, and significantly improved the prediction performance of the model. Signal decomposition techniques can decompose complex surface settlement time-series data into simple subsequences, remove the noise components from the original data, and consequently improve the prediction model's capability to extract internal data features. Common signal decomposition techniques such as EMD (Empirical Mode Decomposition) (Lei et al., 2013) and WT (Wavelet Transform) (Lin and Qu, 2000) are prone to modal aliasing problems when processing nonlinear time-series data (Ge et al., 2025; Gao et al., 2023; Lu et al., 2022). In contrast, VMD (Variational Mode Decomposition) constrains the bandwidth and frequency range of modal functions by introducing a regularization term, thereby effectively reducing modal aliasing problems and achieving better signal decomposition results (Dragomiretskiy et al., 2014). In extracting and recovering bridge temperature data, Huang et al. (2025) compared the performance of VMD, EMD, and CEEMD (Complete Ensemble Empirical Mode Decomposition) in decomposing temperature data, and found that VMD significantly outperformed the other two methods. In diagnosing friction faults in rotor systems, Wang et al. (2015) applied VMD to decompose vibration signals, successfully extracting multiple fault features such as synchronous vibration and subharmonics, and compared the performance of VMD

with EMD, EWT (Empirical Wavelet Transform), and EEMD (Ensemble Empirical Mode Decomposition), demonstrating that VMD significantly outperformed other methods in simultaneously extracting multiple features. By comparing decomposition algorithms including VMD, EMD, EEMD, CEEMDAN (Complete Ensemble Empirical Mode Decomposition With Adaptive Noise), WD (Wavelet Decomposition), and WPD (Wavelet Packet Decomposition), Huang et al. (2021) found that VMD demonstrated superior performance in traffic flow data decomposition, and exhibited the strongest stability in multi-step prediction.

The structural parameters of neural networks directly affect the prediction performance, and improper settings easily lead to non-convergence or fitting problems, so it is necessary to find a set of optimal structural parameters to improve the model's prediction performance. In addition, this paper has chosen 15 feature parameters related to surface settlement. Each of these feature parameters exerts a varying degree of influence on surface settlement. In order to enhance the prediction model's capacity to extract crucial information, it becomes essential to assign distinct weights to these feature parameters when they are input into the BiLSTM. To solve the above problems, a hybrid model (VSA-BiLSTM) integrating VMD, SSA, ATT and BiLSTM is developed in this paper for surface settlement prediction. In contrast to general models that are unable to recognize and deal with noise, this method uses VMD to decompose the surface settlement data into multiple intrinsic modal functions (IMFs) to eliminate the effects of noise. The hyperparameters of BiLSTM are optimized using Sparrow Search Algorithm (SSA) to prevent model overfitting (Liu et al., 2024; Zhou et al., 2024). Attention mechanisms (ATT) are used to prioritize key feature parameters to improve the model's ability to extract key information (Zhou et al., 2022). These methods not only improve the predictive performance of the model, but also address the limitations of previous models in dealing with data abruptness and noise. Finally, the constructed SSA-ATT-BiLSTM model is applied to train and predict each VMD-decomposed subsequence individually, then the prediction results of all subsequences are summed to obtain the final surface settlement prediction. Taking the Shenzhen Dayawan Intercity Tunnel project as an example, the case data is used to validate that the model outperforms existing methods (BiLSTM, SVR, RNN, BPNN), thereby demonstrating the effectiveness of the model proposed in this paper.

The theoretical breakthroughs and engineering value are as follows.

- (1) Methodological Integration Innovation: This paper integrated, for the first time, VMD (for data denoising) from the signal decomposition domain, SSA (for model hyperparameter optimization), BiLSTM (as the main prediction model), and ATT (for feature weighting). Thereby providing a novel VMD-SSA-ATT-BiLSTM collaborative prediction architecture, which addressed the challenges of low data quality and low prediction accuracy in settlement prediction.
- (2) Parameter Optimization Breakthrough: The SSA algorithm employed achieves rapid convergence within 60 epochs, demonstrating superior efficiency compared to traditional optimization algorithms.
- (3) Engineering Applicability: After introducing VMD, the SNR of the data acquired in noisy environments was significantly improved by 7.2–8.5 dB, effectively reducing data noise, which fully demonstrated the benefits of this method in practical applications.

The main contributions of this paper are summarized below.

- (1) Data decomposition stage: The number of modes K is predetermined through the center frequency method, and the VMD approach is employed to decompose the surface settlement data into K more stable subsequences while eliminating the noise effects.

Table 1

Application and comparison of advantages and disadvantages of different machine learning models in the field of prediction in recent years.

Reference	Method	Input	Advantage	Limitation
Shan et al. (2024a)	RNN	Tunneling rate, Geological parameters, Operating parameters	Capable of processing time series data	Gradient vanishing or explosion problems are likely to occur, making it difficult to retain key information in long sequences.
Chen et al. (2024)	AWC-LSTM	Surface deformation time series	Solves the RNN gradient vanishing and explosion problems	Only considering unidirectional information transfer, failing to reflect the interaction patterns of sequential data during tunnel construction
Liu et al. (2023)	GA-VMD-SE-BiLSTM	Wind speed data	Combining forward and backward LSTM, capable of utilizing both past and future information for prediction	When the model structure is not arranged reasonably, the model calculation efficiency may decrease.

- (2) Model optimization stage: SSA is used to optimize the structural parameters of the BiLSTM prediction model (number of training times, batch size, dropout value, number of hidden layer layers, number of hidden units per layer) to improve the model prediction performance.
- (3) Model prediction stage: A hybrid prediction framework combining VMD, SSA, and ATT is proposed, and a BiLSTM model with bidirectional timing information processing capability is utilized as a predictor to improve the hybrid model prediction capability.
- (4) Experimental validation stage: Comparison experiments with BiLSTM, SVR, RNN, BPNN, and a variety of evaluation indicators as the basis of judgment, to verify the superior performance of the proposed model in surface settlement prediction.
- (5) Feature analysis stage: Utilizing SHAP analysis to quantify feature contributions, the influence degree of each feature parameter on surface settlement prediction results is revealed, thereby enhancing the model's interpretability and guiding engineering decision-making.

2. Theoretical basis

The VSA-BiLSTM research method proposed in this paper consists of four main parts. The first part is the VMD for decomposing the time-series signals of surface settlement, the second part is the SSA algorithm for optimizing the hyperparameters of the prediction model, the third part is the BiLSTM model for predicting surface settlement, and the fourth part is the ATT used to assign weights to the features extracted by BiLSTM.

2.1. Variational mode decomposition

VMD (Variational Mode Decomposition) is an adaptive method for decomposing non-smooth signals. The central goal of decomposition is to decompose a complex signal into a specified number of narrowband modal functions and adaptively determine the center frequency of each mode (Liu and Liu, 2023). Applying VMD to signal analysis enables the decomposition of surface settlement data into a specified number of physically meaningful IMFs (Intrinsic Mode Functions), which facilitates the analysis of the local frequency characteristics of the surface settlement data (Zhang et al., 2023). The specific process of VMD decomposition is:

(1) Constructing variational problems

The original signal sequence is decomposed into k IMFs, i.e., modal components (Liu et al., 2020). To ensure the minimized sum of bandwidths for these modal components, the constrained variational formulation of VMD is first defined as:

$$\begin{cases} \min_{\{u_k\}, \{w_k\}} \left\{ \sum_{k=1}^k \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) u_k(t) \right] e^{-jw_k t} \right\|_2^2 \right\} \\ s.t. \sum_{k=1}^k u_k(t) = f(t) \end{cases} \quad (1)$$

where $\{u_k\} = \{u_1, u_2, \dots, u_k\}$ is the k modal components obtained after decomposition; $\{w_k\} = \{w_1, w_2, \dots, w_k\}$ is the center frequency corresponding to each modal component; $f(t)$ is the original time series signal.

(2) Reconstructing the constraint problem

To solve the above problem, Equation (1) transforms the constrained variational problem into a more tractable unconstrained variational

problem by introducing a penalty factor α with the Lagrange multiplier operator λ , whose mathematical expression is:

$$L(\{u_k\}, \{w_k\}, \lambda) = \alpha \sum_{k=1}^k \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) u_k(t) \right] e^{-jw_k t} \right\|_2^2 + \left\| f(t) - \sum_{k=1}^k u_k(t) \right\| + \left\langle \lambda(t), f(t) - \sum_{k=1}^k u_k(t) \right\rangle \quad (2)$$

where λ is the Lagrange operator; α is the quadratic penalty factor.

(3) Solving unconstrained problems

The unconstrained variational problem is solved using the alternating direction method of multipliers, which constantly updates the modal function $u_k(t)$ and center frequency w_k , and finally obtains the optimal solution expression for the modal function and center frequency as:

$$\hat{u}_k^{n+1}(w) = \frac{\hat{f}(w) - \sum_{i=1, i \neq k}^k \hat{u}_i^n(w) + \frac{\lambda}{2}(w)}{1 + 2\alpha(w - w_k^n)^2} \quad (3)$$

$$w_k^{n+1} = \frac{\int_0^\infty w |\hat{u}_k^{n+1}(w)|^2 dw}{\int_0^\infty |\hat{u}_k^{n+1}(w)|^2 dw} \quad (4)$$

(4) Setting the iteration stop condition

The update stops when the precision condition is satisfied, and the iteration stopping expression is:

$$\sum_{k=1}^k \frac{\|\hat{u}_k^{n+1}(w) - \hat{u}_k^n(w)\|_2^2}{\|\hat{u}_k^n(w)\|_2^2} < \varepsilon \quad (5)$$

where ε is the convergence threshold.

The iteration terminates when Eq. (5)'s condition is satisfied, yielding K intrinsic mode functions $\{u_k\}$ with minimized aggregate bandwidth, where the original time-series signal can be reconstructed by summing $\{u_k\}$ with the residual component.

2.2. BiLSTM neural network

LSTM (Long Short-Term Memory) neural networks can be used to process and predict time series data that possess long-term dependencies (Liu and Liang, 2024). Each LSTM cell is composed of 1 memory cell and 3 gating cells (forget gate, input gate, and output gate). The memory cell serves as an information conveyor belt to maintain long-term information, while the gating mechanism selectively retains critical historical information and discards unimportant information, with their combined operation enabling effective capture of long-term dependencies in time series data (Xu et al., 2024; Ye et al., 2024b).

BiLSTM (Bidirectional Long Short-Term Memory) consists of two LSTM layers with opposing information flow directions, enabling the model to simultaneously process temporal data in both forward (input-to-output) and backward (output-to-input) directions (Li et al., 2022). Through this architecture, BiLSTM can obtain more comprehensive contextual information in the whole time series, which overcomes the limitation of LSTM networks in insufficient feature extraction from data (Xu et al., 2023b). The BiLSTM model structure is shown in Fig. 1. The detailed calculation process for LSTM and BiLSTM is provided in Appendix A.

Although the BiLSTM model performs well in the task of time-series data prediction, its accuracy is still affected by the hyperparameters of the model. Therefore, in this paper, the SSA is selected to optimize the

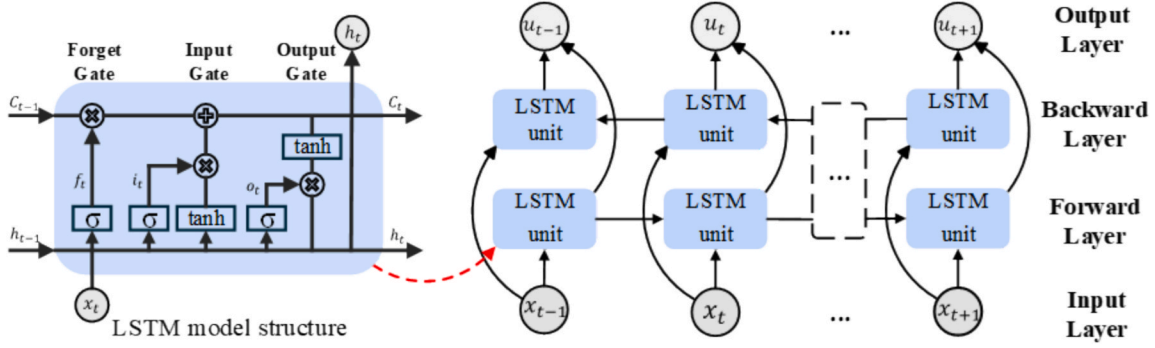


Fig. 1. BiLSTM model structure.

hyperparameters of the BiLSTM model in order to improve the prediction accuracy and prediction efficiency of the model.

2.3. Sparrow search algorithm

The SSA (Sparrow Search Algorithm) is an optimization algorithm that simulates the foraging and anti-predation behaviors of sparrow populations to identify optimal solutions. The sparrow population is categorized into 3 types of roles: discoverer, follower, and vigilant, where each individual sparrow's position represents a solution to the objective function. Through iterative position updates of these three role groups, the optimal solution can be effectively identified (Chen et al., 2022; Lei et al., 2024).

- (1) The discoverer is responsible for finding the optimal solution location and providing the search area and direction to the entire sparrow population. The formula for updating the position of the discoverer is as follows (Zhang et al., 2024).

$$X_{ij}^{t+1} = \begin{cases} X_{ij}^t \cdot \exp\left(\frac{-i}{\alpha i_{\text{termax}}}\right), R_2 < S_T \\ X_{ij}^t + Q \cdot L, R_2 \geq S_T \end{cases} \quad (6)$$

where X_{ij}^{t+1} represents the position of the i -th sparrow in the j -th dimension at the $(t+1)$ -th iteration; α is a random number taking the value $(0,1]$; i_{termax} is the maximum iteration number of the population; Q is a random number obeying a normal distribution; L is a matrix whose every element is 1; R_2 is the alarm value; S_T is the alarm threshold.

- (2) The follower searches for the optimal solution based on the information provided by the discoverer. The follower position update formula is as follows:

$$X_{ij}^{t+1} = \begin{cases} Q \cdot \exp\left(\frac{X_{\text{worst}}^t - X_{ij}^t}{i^2}\right), i > \frac{n}{2} \\ X_p^{t+1} + |X_{ij}^t - X_p^{t+1}| \cdot A^+ \cdot L, i \leq \frac{n}{2} \end{cases} \quad (7)$$

where X_{worst}^t is the global worst individual position in the t -th iteration; X_p^{t+1} is the global best individual position in the $(t+1)$ -th iteration; A is a matrix with the same dimensions as L , where each element is assigned a value of 1 or -1 , $A^+ = A^T(AA^T)^{-1}$.

- (3) The watchers monitor the positions and behaviors of other sparrows to prevent the search process from falling into local optima, if the warning value exceeds the upper limit, the watchers instruct other sparrows to adjust their positions to avoid getting trapped (Xie et al., 2023). The position update formula for the watchers is as follows:

$$X_{ij}^{t+1} = \begin{cases} X_{\text{best}}^t + \beta |X_{ij}^t - X_{\text{best}}^t|, f_i > f_g \\ X_{ij}^t + K \cdot \frac{|X_{ij}^t - X_{\text{worst}}^t|}{(f_i - f_w) + \epsilon}, f_i = f_g \\ X_{ij}^t, f_i < f_g \end{cases} \quad (8)$$

2.4. Attention mechanisms

The attention mechanism is a method that enables models to focus on important information while ignoring irrelevant data during processing. This mechanism prevents the problem of losing critical information due to long training data sequences. The attention mechanism computes the similarity between the query vector and each key vector to obtain their corresponding value vectors, then performs weighted summation on these value vectors to generate the final attention value, where distinct attention values represent the weights assigned to different features (Yu et al., 2024; Kang et al., 2023). The process of calculating the attention mechanism is shown in Fig. 2, and the calculation formula is as follows (Zou et al., 2024):

$$Q = X \cdot W_Q, K = X \cdot W_K, V = X \cdot W_V \quad (9)$$

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{Q \cdot K^T}{\sqrt{d_K}}\right) V \quad (10)$$

where Q , K , and V are query vector, key vector, and value vector respectively; W_Q , W_K , and W_V denote the corresponding weight matrices of Q , K , and V vectors respectively. and d_K is the dimension of the key vector.

Combining all of the above modules, the overall architecture of the VSA-BiLSTM model was constructed, and its structural diagram is shown in Fig. 3.

3. Case study and data acquisition

The study project is the interval from Minzhi North Station to Longsheng Station of Shenzhen Daya Bay Intercity Tunnel. The tunnel is constructed beneath Bulong Road, successively under-crossing high-rise buildings, Metro Line 6, highway bridges, underground passages, and metro stations, with a total length of 2.35 km. The depth of the shield tunnel is 15–24 m, the outer diameter of the tunnel is 4.4 m, the inner diameter is 4 m, the ring width is 1.8 m, and the maximum longitudinal slope of the line is 4.8 %. The engineering location and shield route map are shown in Figs. 4(a) and (b). The tunnel was constructed using a ZTE9100 composite earth pressure balance shield machine. The shield tunnel and shield machine are shown in Fig. 4(c).

The study interval tunnel is located in a densely populated area of the city, and there are many obstacles along the shield tunneling route, with the distribution of viaduct pile foundations, pile foundations of medium-

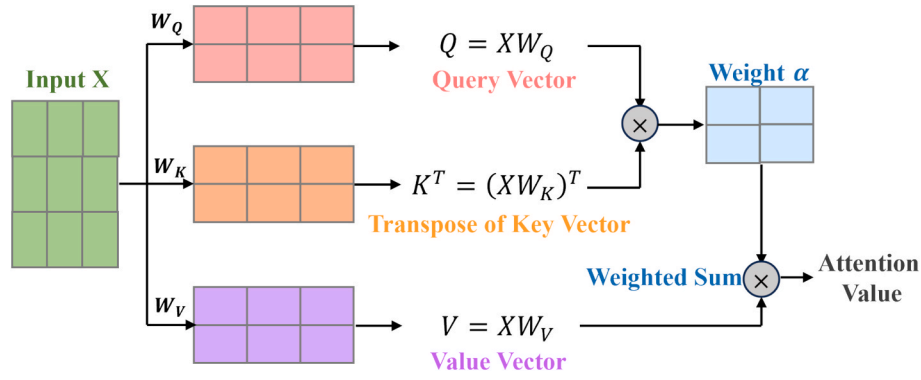


Fig. 2. Attention mechanism calculation flow diagram.

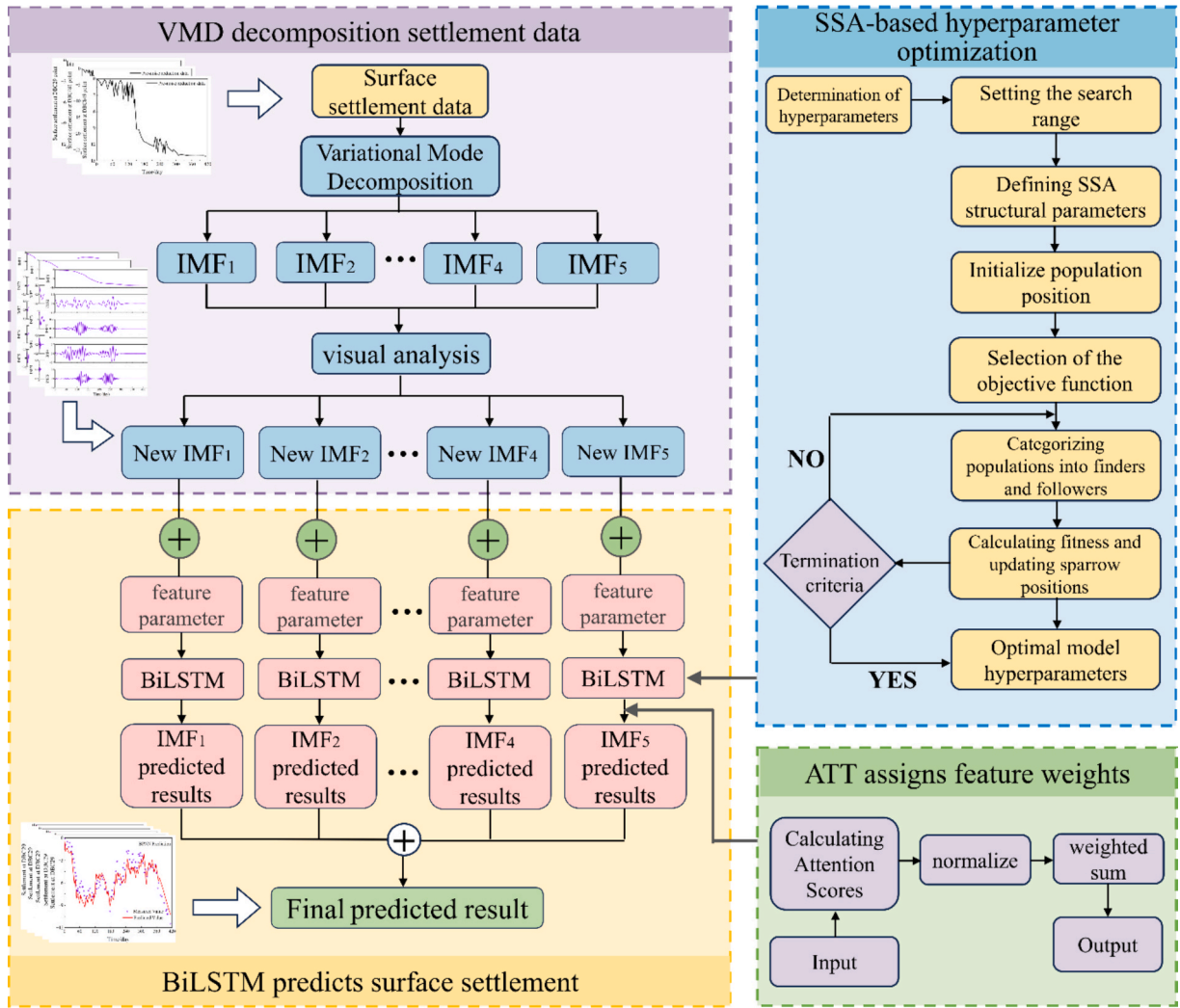


Fig. 3. VSA-BiLSTM model analysis process.

and high-rise buildings, and other existing structures, among which the minimum clearance between the bottom of pedestrian bridge piles and the upper edge of the tunnel is only 0.77 m. Groundwater is relatively abundant along the line, mainly in the sand layer, with the water content of the completely weathered granite layer and the highly weathered granite layer being the next highest. These aquifers are distributed along the main shield tunneling route and are prone to induce relatively significant surface settlement during the excavation process. To ensure

tunnel construction safety, monitoring points were installed along the tunnel axis with a spacing interval of 10 m. The monitoring frequency for each point was 1 day during the initial phase, then adjusted to 15 days after data stabilization, with missing data filled using nonlinear interpolation. The monitoring period of the monitoring points is from November 16, 2023, to December 31, 2024, with a total of 412 days of data, from which the maximum surface settlement value during shield tunneling is obtained and used as the output parameter of the dataset in

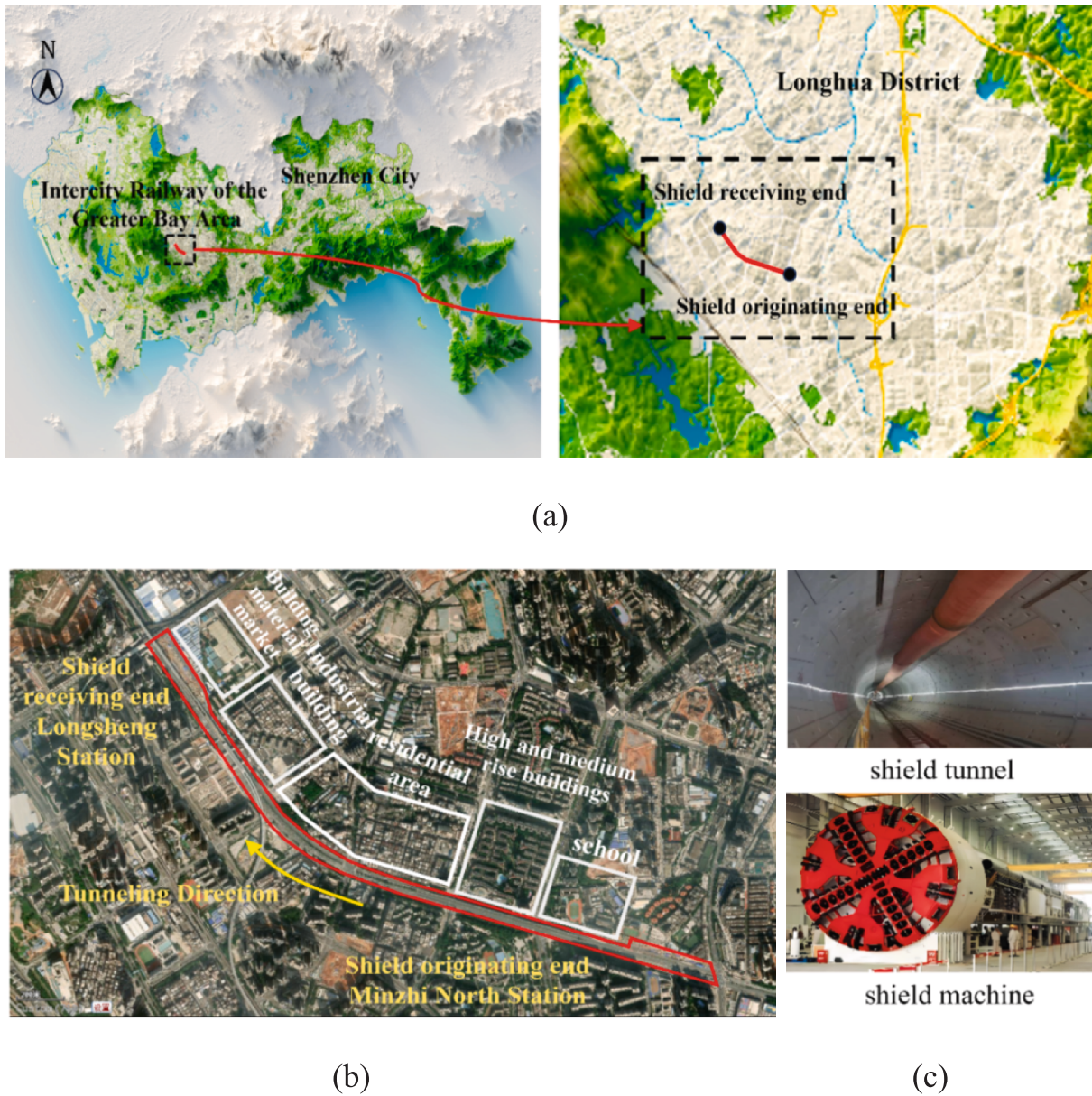


Fig. 4. Engineering location map: (a) Shenzhen engineering location map. (b) Engineering interval route detailed map. (c) Shield tunnel and shield machine.

this paper.

When analyzing the surface settlement data generated by shield tunneling, it was found that when the shield machine first passed through the tunnel below the monitoring points, three monitoring points recorded significantly higher surface settlement values than other monitoring points in the previous seven days. It can be seen that the cumulative settlement of these three points is greater, so these three monitoring points were studied. The three monitoring points are M29-L, M481-L and M849-L. Fig. 5 shows the surface settlement changes recorded at these three monitoring points during the first 7 days under the initial disturbance of the shield machine. As can be seen from the figure, the cumulative settlement of the three monitoring points reached 2.95 mm, 3.24 mm, and 2.67 mm respectively on the seventh day, and the surface settlement curve showed a continuous downward trend, so it is significant to study the subsequent settlement of these three points. The location maps of the three monitoring points are shown in Fig. 6 (a). As can be seen from the figure, these three monitoring points are all located above the left tunnel, mainly because of the special surrounding environment in which they are located, M29-L is near the school area,

M481-L is close to the high-rise building, and there is a cross-tunnel underneath M849-L, and these complex conditions lead to larger surface settlement occurring in the left tunnel.

The dataset constructed in this paper includes input feature parameters and output parameters, and the input feature parameters include three categories: geological parameters, geometric parameters, and shield machine parameters. The geologic profile and tunnel locations within the study area are shown in Fig. 6 (b). The soil layers within the shield crossing area mainly include ① plain fill, ② clayey soil, ③ gravelly clayey soil, ④ completely weathered granite, ⑤ soil-like highly weathered granite, ⑥ moderately weathered granite, and ⑦ slightly weathered granite. The input feature parameters (geological parameters, geometric parameters, shield machine parameters) in the dataset are shown in Table 2. The selection of these input feature parameters is based on previous research and domain expertise (Liu et al., 2022; Wang et al., 2025), which indicates that the selected feature parameters influence the interaction between the soil and the tunnel structure during tunnel boring. The output parameters selected for this paper are surface settlement data for monitoring points M29-L, M481-L, and M849-L. In

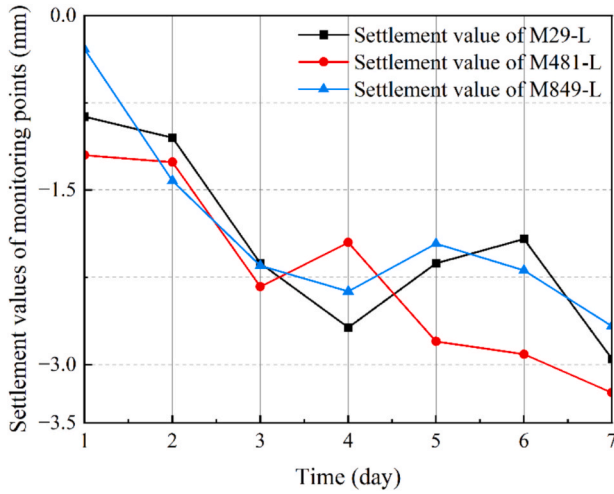


Fig. 5. Surface settlement curves at the three monitoring points during the initial period of shield disturbance.

the following sections, these three monitoring points are abbreviated as ML29, ML481, and ML849.

The dataset constructed in this paper contains three monitoring points, each monitoring point contains 412 samples of surface settlement data, each sample includes 15 input feature parameters and 1 output parameter. Since the surface settlement data in this paper are derived from continuous monitoring, it exhibits strong time dependence. To better preserve the temporal dependence of the data, we chose to divide the dataset chronologically. Specifically, the first 80 % (330 days) of the dataset was used as the training set, and the remaining 20 % (82 days) was used as the test set (Qin et al., 2024).

4. Establishment of surface settlement prediction model

4.1. Data preprocessing

To improve the training efficiency of the prediction model, the input feature data (shield machine parameters, geological parameters, geometric parameters) undergo data preprocessing, including missing value filling and normalization. For the missing values accounting for less than 1 % of the data set, this paper uses a nonlinear interpolation method to supplement them. Since some feature data values range from 0 – 0.5, and some data ranges from 2500 – 4500, normalization is applied to the feature data. The normalization method uses the min–max approach, scaling each feature data to the range [0,1], with the formula expressed as:

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (11)$$

where x' is the normalized data; $x_{\max}$ is the maximum value in the dataset; $x_{\min}$ is the minimum value in the dataset.

4.2. VMD configuration

VMD will decompose the surface settlement data into K intrinsic mode functions. In order to fully decompose the surface settlement data, a reasonable number of mode decompositions K needs to be selected. An excessively small K value leads to an under-decomposition of data, while an overly large K value results in modal repetition, both scenarios adversely affecting data analysis (Gong et al., 2019; Gan et al., 2025). Therefore, in this paper, the center frequency method is used to determine the value of K (Wen et al., 2024; Fu et al., 2019). When the center frequency of the last layer of components tends to be stable, it means that the decomposition of the data has been more thorough. The number

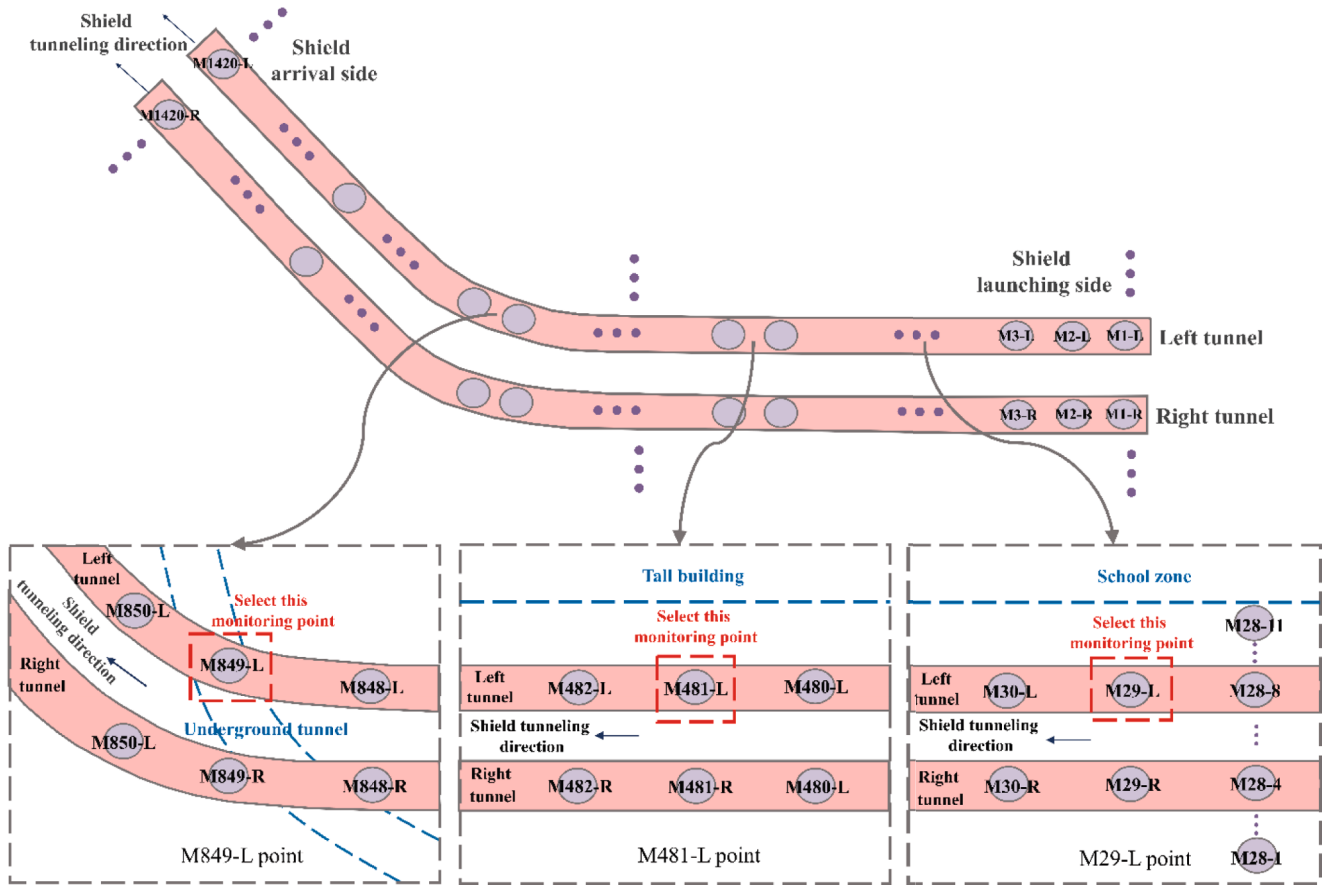
of decomposed modes is set to $K = 1, 2, 3, \dots, n$ respectively, and the variational mode decomposition is performed for each case, and the center frequency of the mode components is obtained for each case. This paper investigates the surface settlement data at three monitoring points during shield tunneling advancement, with the three monitoring points being ML29, ML481, and ML849 monitoring points. Taking the surface settlement data at monitoring point ML29 as an example, the number of VMD modal decompositions is determined.

When K takes values from 1 to 7 respectively, the center frequencies of the respective IMFs obtained after VMD decomposition are shown in Table 9 of Appendix B. As shown in Table 9, when $K = 7$, two similar frequencies appear: $\text{IMF6} = 0.277$ and $\text{IMF7} = 0.324$. It can be concluded that when $K = 7$, the data begins to be over-decomposed, manifesting as mode mixing, where frequency components overlap between different modes. As indicated by the footnote formula¹, when the sum of the absolute differences between the center frequencies of two adjacent IMFs divided by their average center frequency is less than the threshold μ , the decomposition is considered sufficiently refined, and further decomposition is stopped. When $K = 7$, the center frequencies of IMF7 and IMF6 are extremely close, and the minimum variation of the center frequency for IMF7 within the adjacent frequency domain is 0.0156, satisfying the requirement in the footnote formula that the decomposition threshold should be less than 0.2. The Fig. 7 shows the RMSE, MAE, and R^2 obtained from the BiLSTM predictions for the ML29 monitoring point under different modal component K values. As can be seen from the figure, as the number of modes K increased, the RMSE and MAE generally showed a trend of first decreasing and then increasing. When $K = 6$, the RMSE and MAE achieve their minimum values. R^2 showed a trend of first increasing and then decreasing. Although R^2 had a value of 0.873 when K was 6, it reached its highest value of 0.874 when K was 7. but when all indicators were considered comprehensively, the model achieved its best prediction performance when K was 6. Once again, when $K = 7$, modal decomposition becomes excessive, so the number of modal decompositions K is selected as 6. Actively discard the noise component IMF6 in the modal number $K = 6$, select data IMF1 to IMF5 to reconstruct the surface settlement data, namely the denoised data. The comparison of surface settlement data before and after denoising for each monitoring point, along with the modal decomposition results, is shown in Fig. 8. The ML481 and ML849 monitoring points also use the center frequency method to determine the K value. In the VMD algorithm, VMD decomposition is implemented using the Python third-party library vmdpy. Other key parameters such as α (Penalty Factor) = 2000, tol (Convergence/Decomposition Tolerance) = $1e-7$, and τ (Noise Tolerance) = 0.1.

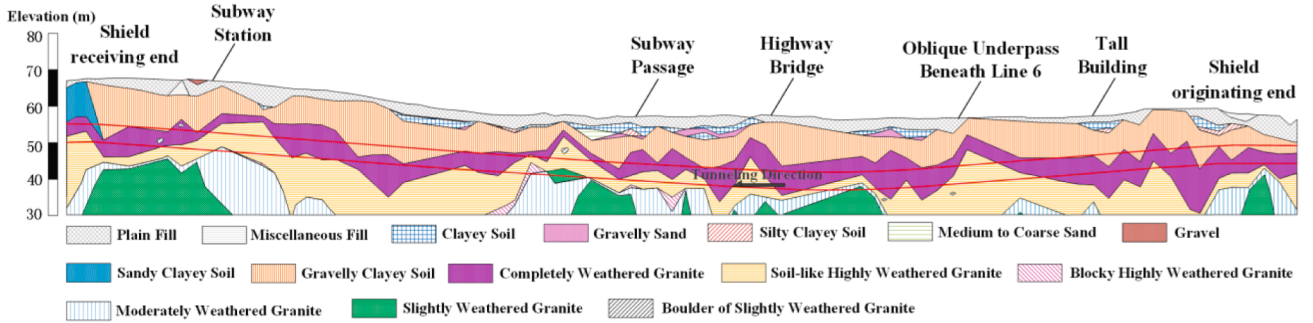
SNR (Signal-to-Noise Ratio) is the ratio of signal power to noise power and is an important parameter for measuring signal quality. A higher SNR represents less noise interference and better signal quality, conversely, a lower SNR represents more noise interference and poorer signal quality. While VMD effectively reduced noise, which was specifically manifested as a SNR improvement of 7.2–8.5 dB. The calculation method for SNR is given in Eq. (12). According to the formula calculation, the SNR of the original surface settlement data at the ML29 monitoring point is 5.3 dB, and after the VMD decomposition, the IMF6 is judged as noise and removed, and the remaining IMF reconstructed noise-reduced signal has an SNR of 12.5 dB, which is an overall improvement of 7.2 dB compared with the pre-noise reduction. The SNR of the surface settlement data at monitoring point ML481 improved from 4.6 dB to 13.1 dB, an improvement of 8.5 dB. The SNR for surface settlement data at monitoring point ML849 increased from 6.5 dB to 13.9 dB, an improvement of 7.4 dB. These results show that VMD can effectively suppress noise and improve signal quality.

¹ $\min \left(\left| \frac{w_{i+1} - w_i}{(w_{i+1} + w_i)/2} \right| \right) <$

μ Where μ is the decomposition threshold and is taken as 0.2.



(a)



(b)

Fig. 6. Comprehensive Engineering-Geological Diagram: (a) Location map of the three monitoring points (b) Geological profile of shield tunneling section.

$$SNR_{dB} = 10 \cdot \log_{10} \left(\frac{P_{signal}}{P_{noise}} \right) \quad (12)$$

where P_{signal} is the power of signal; P_{noise} is the power of noise.

4.3. SSA hyperparameter optimization

To improve the prediction accuracy of the model, using RMSE as the objective function and Adam as the optimizer, the sparrow search algorithm is employed to optimize the hyperparameters of the BiLSTM. In this paper, five hyperparameters that have a large impact on the model prediction performance are selected for optimization, including the

number of training times, batch size, dropout value, the number of hidden layer layers, and the number of hidden units per layer. To improve search efficiency and reduce computing costs, the search range for the number of hidden layers is set to 2–6 layers, and the search range for the number of hidden layer units is set to 10–120; The search range for the number of training times is set to 10–150; The search ranges for batch size and dropout value are set to 20–150 and 0.1–0.5, respectively.

The real values of surface settlement at monitoring point ML481 were used as input for the hyperparameter optimization of the model. Fig. 9 demonstrates the RMSE values and total training time of the model under different training epochs. It can be seen that the total training time curve shows a continuous upward trend as the number of training

Table 2
Feature parameter table.

Categorization of feature parameter	Feature parameter Codes	physical meaning	unit
geological parameter	X1	Deformation modulus of gravelly clayey soil layer at the top of the tunnel	kPa
	X2	Cohesion of gravelly clayey soil layer at the top of the tunnel	
	X3	Poisson's ratio of gravelly clayey soil layer at the top of the tunnel	
	X4	Deformation modulus of soil-like highly weathered granite layer at the bottom of the tunnel	kPa
	X5	Cohesion of soil-like highly weathered granite layer at the bottom of the tunnel	
	X6	Poisson's ratio of soil-like highly weathered granite layer at the bottom of the tunnel	
geometric parameter shield machine parameter	X7	Average tunnel depth	m
	X8	tunnel diameter	
	X9	cutterhead rotation speed	rmp
	X10	cutterhead torque	KN•m
	X11	grouting volume	L
	X12	grouting pressure	bar
	X13	shield machine pitch angle	°
	X14	Maximum digging speed	mm/min
	X15	total shield thrust	KN

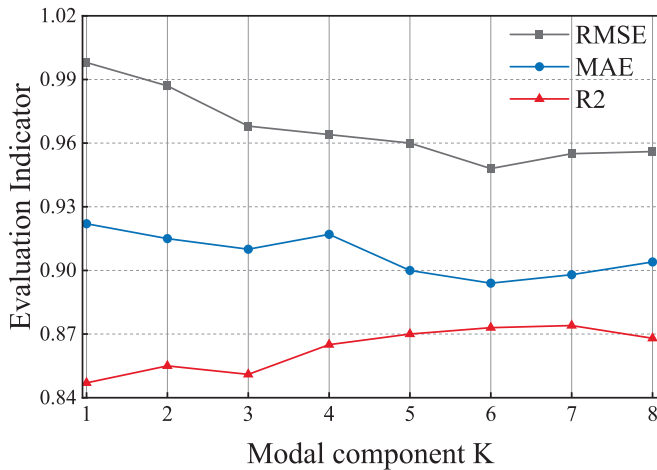


Fig. 7. Evaluation indicators for BiLSTM under different modality components.

epochs increases. The RMSE curve showed a decreasing trend during 0–60 training epochs, reaching its lowest value of 2.305 at 60 epochs. The RMSE curve showed an upward trend between 60–102 epochs, then decreased significantly in the 102–140 epoch range, reaching its minimum value of 2.303 at 140 epochs. The RMSE values at 60 and 140 epochs were comparable. Although the root mean square error at epoch 140 was slightly lower (2.303 vs 2.305), epoch 60 was ultimately selected based on time cost considerations. Compared with other traditional optimization algorithms (e.g., PSO, GA), SSA requires the fewest iterations to find feasible solutions and converges the fastest, demonstrating its superior optimization efficiency among these algorithms (Shen et al., 2025). This method was used to continue optimizing other hyperparameters, and the results of the model hyperparameter optimization are shown in Table 3. The key hyperparameter settings for the SSA algorithm are shown in Table 10 of Appendix B. The pseudocode for the VMD-SSA module is provided in Appendix C.

4.4. VSA-BiLSTM architecture

The VSA-BiLSTM model proposed in this paper is constructed based on the optimized parameters from the previous section, and the specific workflow for its surface settlement prediction process is shown in Fig. 10.

VMD can be used to denoise the data and extract its main features, thereby improving the data quality; The BiLSTM model can capture the forward and backward dependencies in the data, allowing comprehensive learning of the patterns between data features; The SSA algorithm is used to optimize the hyperparameters of the BiLSTM model, establishing a prediction model with optimal parameter combinations; Finally, ATT assigns corresponding weights to the outputs of the BiLSTM model and performs weighted summation, which makes the model focus on important information and reduces the impact of redundant information on the prediction results.

After adding ATT into the BiLSTM model, the attention score is first computed to measure the importance of each time step, and in the process, the hidden state of BiLSTM is nonlinearly transformed using the tanh activation function. The weights (summing to 1) are then obtained by normalizing the attention scores with the softmax function. Then, a weighted summation is performed on the hidden states of BiLSTM to obtain a feature vector that can focus on key surface settlement time steps. Finally, this feature vector is input into the fully connected layer to get the final surface settlement value.

The following is the workflow of the VSA-BiLSTM model:

- (1) The selected shield machine parameters, geological parameters, geometric parameters, and surface settlement data during shield tunnel boring are taken as model inputs.
- (2) The input feature data (shield machine parameters, geological parameters, geometric parameters) undergoes data preprocessing.
- (3) The surface settlement data were decomposed using VMD to obtain several IMF and residual components, and the residual components were removed.
- (4) After feeding each IMF into the SSA-BiLSTM-ATT prediction model separately, multiple independent prediction models are obtained.
- (5) Finally, the predictions of all the prediction models were summed to obtain the final prediction of surface settlement.

4.5. Model evaluation indicators

To evaluate the prediction accuracy of the proposed VSA-BiLSTM model and comparative models, five evaluation indicators are selected for model assessment. The evaluation indicators were MAE (Mean Absolute Error), RMSE (Root Mean Square Error), MAPE (Mean Absolute Percentage Error), R^2 (Goodness of Fit), and RPD (Relative Percentage Deviation), their calculation is shown in Eqs. 13–17. The lower the MAE, RMSE, and MAPE values, the higher the prediction accuracy; whereas the higher the R^2 and RPD values, the better the fit between predicted and actual values, indicating greater prediction accuracy. In addition, the prediction results of the five models were analyzed using Taylor diagrams, which clearly demonstrate the performance of each model through SD (Standard Deviation), R (Correlation Coefficient), and CRMSD (Central Root Mean Square Deviation), their calculation is shown in Eqs. (18)–(20). The formulas for each evaluation indicator are (Ye et al., 2024a):

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (13)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (14)$$

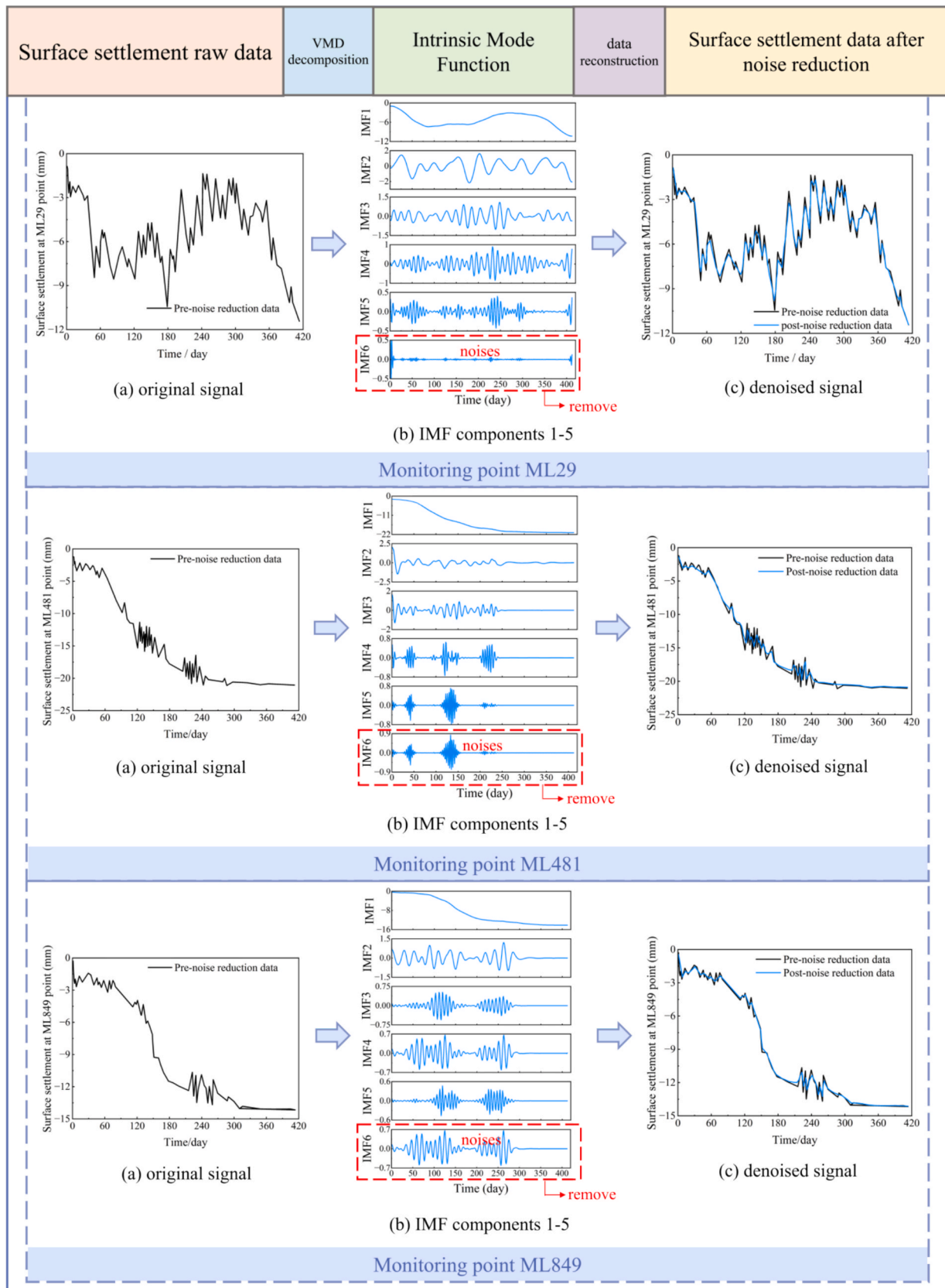


Fig. 8. Denoising Results of Surface Settlement Data from Each Monitoring Point.

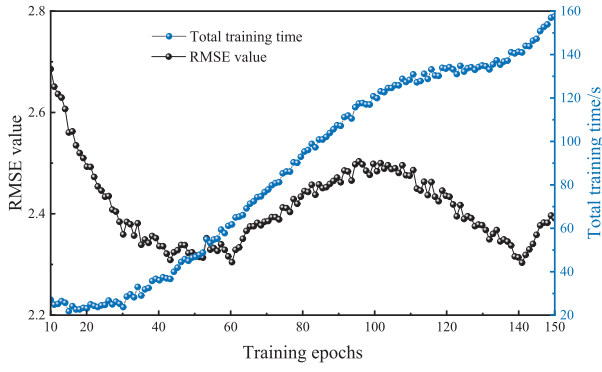


Fig. 9. RMSE and total time consumption under different training epochs.

Table 3
BiLSTM Sparrow Search optimized results.

BiLSTM hyperparameters	Search Range	value
training times	10–150	60
batch size	20–150	64
dropout value	0.1–0.5	0.2
Number of hidden layers	2–6	2
Number of hidden layer units	10–120	64

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100\% \quad (15)$$

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

$$RPD = \frac{SD}{E_{RMSE}} = \frac{\sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2}}{E_{RMSE}} \quad (17)$$

$$R = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (18)$$

$$SD = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2} \quad (19)$$

$$CRMSD = \sqrt{\frac{\sum_{i=1}^n ((\hat{y}_i - \bar{y}) - (y_i - \bar{y}))^2}{n}} \quad (20)$$

where n is the number of predicted samples; y_i is the true value of the i -th sample; $\hat{y}_i$ is the predicted value of the i -th sample; and $\bar{y}$ is the mean value of the i -th sample.

5. Analysis of forecast results

To validate the effectiveness of the VSA-BiLSTM model proposed in this paper, BiLSTM, SVR, RNN, and BPNN models are selected for comparative analysis. SVR (Support Vector Regression) is a regression analysis method based on statistical learning theory, which achieves effective fitting and prediction of data by finding the optimal feature space function that satisfies the ϵ -insensitive loss condition. SVR is highly generalizable and suitable for small sample data. RNN (Recurrent Neural Network) is a type of neural network specifically designed for processing sequential data, consisting of an input layer, hidden layer, and output layer. The recurrent structure embedded in RNN's hidden

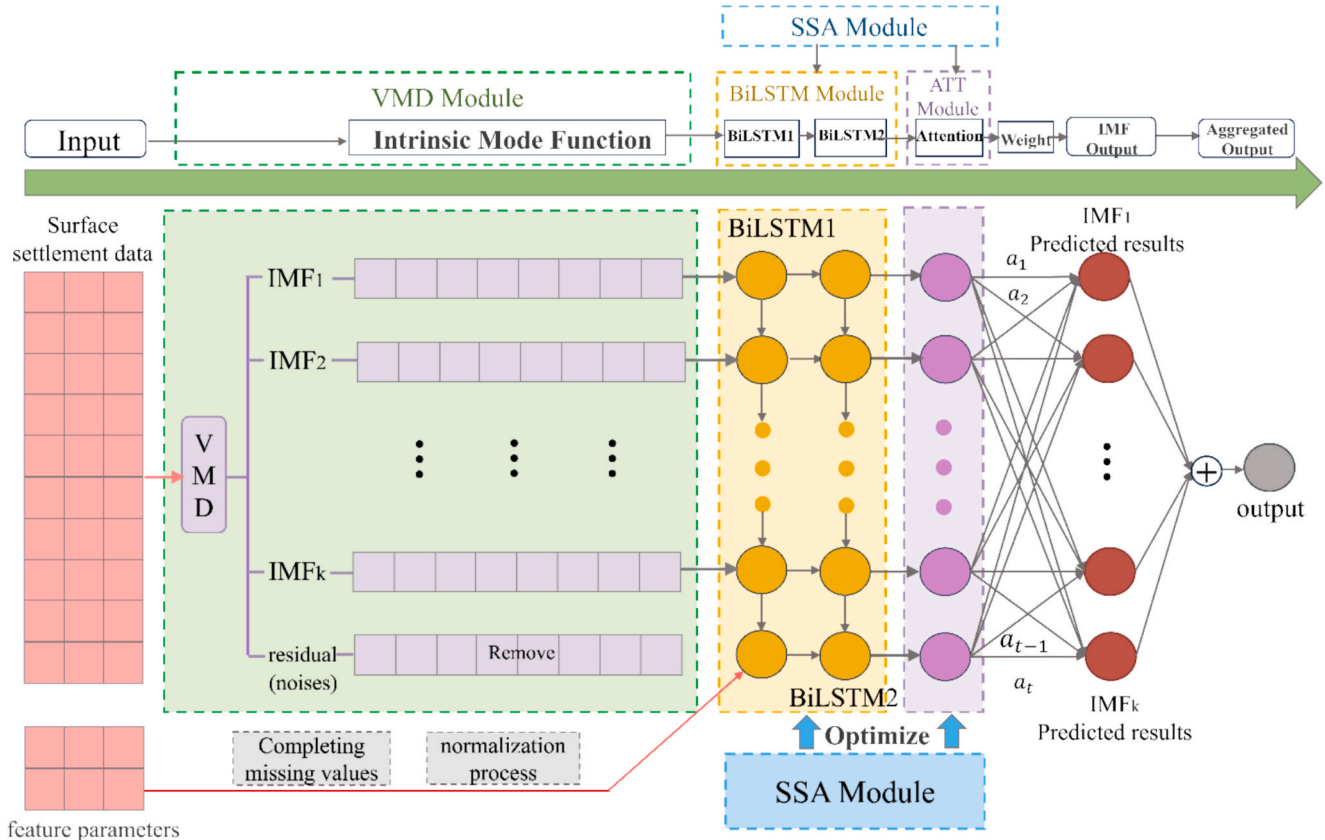


Fig. 10. Flowchart of VSA-BiLSTM model for surface settlement prediction.

layer takes the previous hidden state as input and combines it with the current input to compute a new hidden state, a characteristic that enables RNN to effectively process sequential data. BPNN (Back-propagation Neural Network) is a multilayer feedforward neural network consisting of input, hidden, and output layers. The BPNN forward propagates the output and compares it with the target value to generate the error, and then updates the weights of the neurons by back-propagating the error, so that the loss function is minimized and the output is constantly close to the target value. The specific principles of BiLSTM are described in Section 2.2.

5.1. Comparative model performance

As can be seen from Fig. 11, the surface settlement prediction curves of the five models at the three monitoring points are generally similar to the trend of the measured values, indicating that the hyperparameters configured for each model by the optimization algorithm are more reasonable. Among them, the VSA-BiLSTM model had the highest fit of predicted settlement curves to measured data at the three monitoring points; The settlement prediction curves of both BiLSTM and SVR models at the three monitoring points were generally well fitted to the measured values; The RNN and BPNN models show progressively declining fitting accuracy, with the BPNN model exhibiting the poorest fit to measured values, capable of predicting only their overall trend; In the settlement curve of ML29, there are two continuous oscillation intervals of 60–120 days and 240–300 days, the RNN and BPNN models fit the continuous oscillation intervals poorly compared with the VSA-BiLSTM, BiLSTM, and SVR models, and this phenomenon is also observed at monitoring points ML481 and ML849.

Taylor diagrams have become an increasingly popular graphical tool for evaluating model performance in recent years, offering superior inter-model comparison capabilities compared to single model evaluation methods (Zhao et al., 2023; Su et al., 2024). The VSA-BiLSTM, BiLSTM, SVR, RNN, and BPNN are named as Model 1, Model 2, Model 3, Model 4, and Model 5 in the next analysis.

Fig. 12 and Table 4 show the results of surface settlement predictions for the five models at different monitoring sites. For monitoring point ML29, the positions of the five models in the Taylor diagram are close to the constant standard deviation curve, indicating that the standard deviation difference between the predicted value and the real value is small, and each prediction model can simulate the fluctuation characteristics of the real value better. Among the models, Model 1 demonstrates the smallest difference between the standard deviation of its predicted values (2.135 mm) and that of the actual values (2.524 mm), while Model 5 shows the largest difference (3.076 mm vs 2.524 mm), indicating that Model 1 is better than the other four models in simulating the fluctuating characteristics of the true value. The location of Model 1 on the Taylor diagram is closest to the reference point, indicating that Model 1 has the lowest CRMSD value (0.904 mm) and the highest R value (0.935), which means that Model 1 has the smallest deviation between predicted values and true values and the strongest linear correlation. The remaining four models all have lower R-values than Model 1 and are located roughly along the same correlation coefficient line, with their R-values being relatively similar. Other monitoring points were analyzed in the same way.

From Table 4, it can be seen that Model 1 (VSA-BiLSTM) outperforms the other compared models in prediction performance: its mean SD (Standard Deviation) of predicted values of the three monitoring points

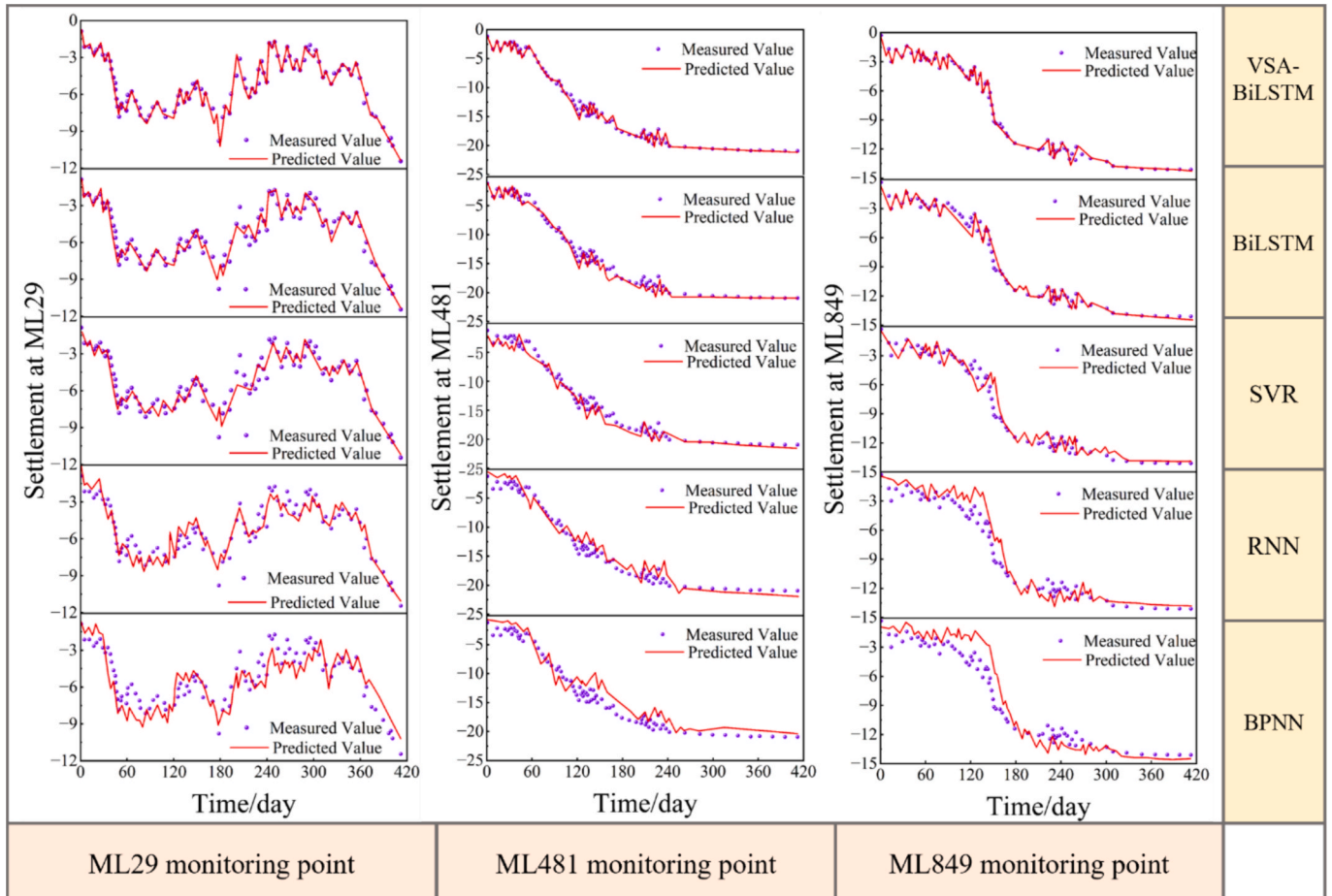


Fig. 11. Comparison of predicted and measured values from five models at three monitoring points.

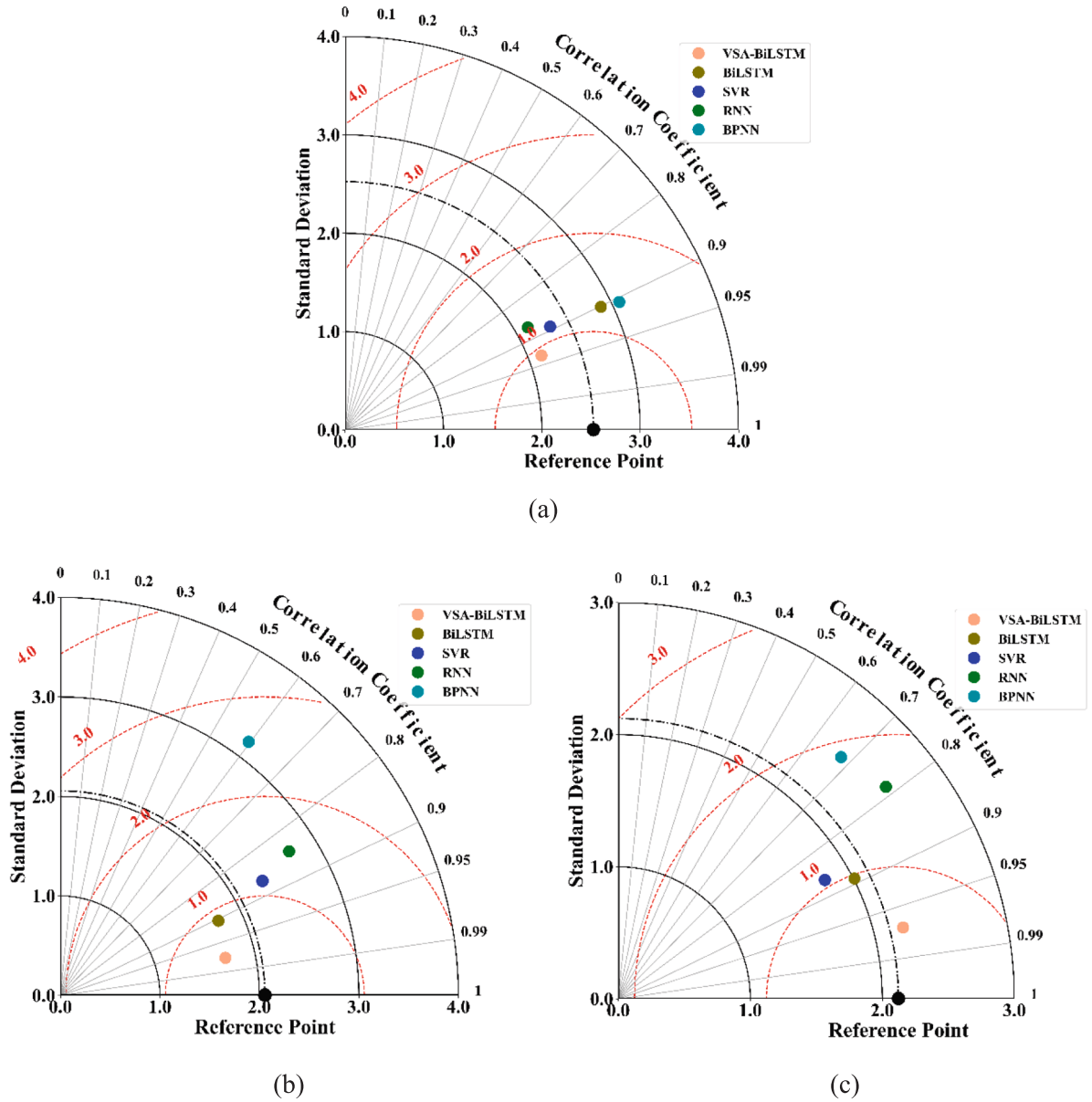


Fig. 12. Taylor diagram of prediction results at three monitoring points: (a) ML29 monitoring point. (b) ML481 monitoring point. (c) ML849 monitoring point.

is 6.39 %–30.66 % lower than other models, the mean R (Correlation Coefficient) is 6.67 %–31.91 % higher, and the mean CRMSD (Centered Root Mean Square Difference) is reduced by 28 %–64.7 % compared to other models. A comprehensive analysis of these three indicators confirms that the VSA-BiLSTM model achieves higher prediction accuracy.

To further explore the prediction performance of the five models, this paper employs five evaluation indicators (RMSE, MAE, MAPE, R^2 , and RPD) to comprehensively assess the models' surface settlement prediction performance at different monitoring points. Taking monitoring point ML481 as an example, Fig. 13 demonstrates the prediction performance of the five models at monitoring point ML481. As can be seen from Fig. 13, model 4 (RNN) and model 5 (BPNN) have overall higher MAE, RMSE, and MAPE values compared to the other models, while the RPD and R^2 values are lower, and the distributions of the two models on the radar plot are mainly concentrated on the right side, which indicates that the predictive performances of the two models are poorer compared to the other models. Model 1 (VSA-BiLSTM) and Model 2 (BiLSTM) exhibit lower MAE, RMSE, and MAPE values than the other models, along with higher RPD and R^2 . Their radar plot distributions are mainly

concentrated in the lower left, indicating that the two models have better predictive performances than the other models. Among them, Model 1 performs more outstandingly, having higher RPD and R^2 values as well as lower MAE, RMSE, and MAPE values compared to Model 2, which is the best performing model among all models.

Table 5 demonstrates the mean evaluation indicators for the different models. The results show that the mean predictive performance of Model 5 is significantly lower than the other four models, with mean RMSE, MAE, MAPE, R^2 and RPD values of 2.516 mm, 1.768 mm, 62.44 %, 0.63, and 1.29, respectively. Model 4, Model 3, and Model 2 exhibit sequentially increasing predictive performance compared to Model 5. Model 1 shows the optimal predictive performance among all models, with mean RMSE, MAE, MAPE, R^2 and RPD values of 1.117 mm, 0.74 mm, 24.24 %, 0.95, and 2.97, respectively. The prediction performance of Model 1 (VSA-BiLSTM) is better than that of Model 2 (BiLSTM), and compared to Model 2 its mean RMSE, MAE, and MAPE are reduced by 0.543 mm, 0.418 mm, and 14.18 % respectively, and the mean R^2 and RPD values are improved by 0.09 and 0.54 respectively, which shows that the introduction of VMD, SSA, and ATT can improve the model's

Table 4
Taylor diagram indicator for three monitoring points.

Monitoring point	Model	Predicted value SD/mm	True value SD/mm	R	CRMSD/mm
ML29	1	2.135	2.524	0.935	0.904
	2	2.886	2.524	0.901	0.998
	3	2.334	2.524	0.893	1.139
	4	2.13	2.524	0.872	1.239
	5	3.076	2.524	0.906	1.329
ML481	1	1.701	2.056	0.975	0.558
	2	1.756	2.056	0.904	0.895
	3	2.334	2.056	0.87	1.208
	4	2.717	2.056	0.846	1.589
	5	3.174	2.056	0.596	2.673
ML849	1	2.224	2.122	0.967	0.606
	2	2.007	2.122	0.891	0.977
	3	1.805	2.122	0.867	1.032
	4	2.586	2.122	0.784	1.651
	5	2.489	2.122	0.678	1.854
mean value	1	2.02	2.234	0.959	0.689
	2	2.216	2.234	0.899	0.957
	3	2.158	2.234	0.877	1.126
	4	2.478	2.234	0.834	1.493
	5	2.913	2.234	0.727	1.952

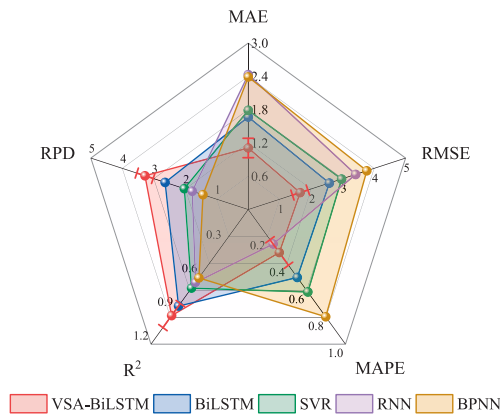


Fig. 13. ML481 Monitoring Point Evaluation Indicators.

Table 5
Mean evaluation indicators for the five models at each monitoring points.

Model	Mean RMSE /mm	Mean MAE /mm	Mean MAPE /%	Mean R²	Mean RPD
1	1.117	0.74	24.24	0.95	2.97
2	1.66	1.158	38.42	0.86	2.43
3	1.9	1.274	45.04	0.77	1.99
4	2.252	1.694	40.4	0.71	1.41
5	2.516	1.768	62.44	0.63	1.29

prediction effect. Due to the presence of zero values in the true values of surface settlement (accounting for less than 1 % of the dataset), which can amplify the MAPE, these zero values are replaced with an extremely small positive number to eliminate their adverse impact on the MAPE.

In summary, in this engineering example, Model 1 (VSA-BiLSTM) has the best prediction performance, while Model 5 (BPNN) showed relatively the poorest predictive capability among the compared models.

Computational efficiency is one of the key indicators for evaluating the performance of a model, and in this study, the prediction time of each prediction method was analyzed, and the results were recorded as shown in Table 6. All prediction models were run on a laptop with a 13th generation Intel Core i7-13700 K CPU, NVIDIA GeForce RTX 4090 GPU, and an operating system of Ubuntu 22.04. As can be seen from the table, the shortest running time on both different datasets is for the single

Table 6
Comparison of prediction time of each prediction model on different surface settlement datasets.

Model	ML29(s)	ML481(s)	ML849(s)
VSA-BiLSTM	5.179	3.079	4.622
BiLSTM	4.565	2.541	3.285
SVR	4.945	2.197	3.045
RNN	15.662	10.386	12.174
BPNN	20.387	16.266	17.085

models BiLSTM and SVR (running time 2–5 s), while the hybrid model proposed in this paper has a slightly longer computation time (running time 3–6 s). The RNN and BPNN models require the longest training time. the proposed hybrid model need longer runtime because it requires decomposing the raw ground settlement data, then training separate BiLSTM models on the decomposed subsequences, thereby introducing additional computational time. although the training time is slightly longer, the proposed hybrid model has the best prediction performance, compared to the 1–2 s time difference with single models, its computational efficiency is still within an acceptable range. Moreover, with the continuous development of computer hardware performance, the running time of the proposed method will be further reduced in the future.

5.2. Robustness verification

To verify whether the error differences between Model 1 (VSA-BiLSTM) and other comparative models are statistically significant, this study employs paired t-tests and Wilcoxon signed rank tests to analyze the RMSE differences among the models. The paired *t*-test was applied when the sample-level error differences between Model 1 and comparative models (Models 2–5) followed a normal distribution, and Wilcoxon signed rank tests were used when they did not. The Shapiro-Wilk test showed that all groups of differences did not fit the normal distribution, so the Wilcoxon signed rank test was chosen. As can be seen from Table 7, all comparisons between the models meet the statistical significance criterion of $p < 0.05$. The RMSE values between Model 1 and Model 2 show a statistically significant difference ($p = 0.0312 < 0.05$), with an effect size Cohen's $d = 0.391$ (95 % CI [0.284, 0.492]), indicating that the difference between the two models reaches a medium level. There is a statistically significant difference between Model 1 and Models 3–4 and this difference increases sequentially. A statistically significant difference was found in RMSE values between Model 1 and Model 5 ($p = 0.0213 < 0.05$), with an effect size Cohen's $d = 0.688$ (95 % CI [0.38, 0.991]), indicating the difference between the two models reached a medium-to-large effect level. The model was also compared with the popular Transformer framework, revealing a statistically significant difference in RMSE values between Model 1 and Transformer ($p = 0.0414 < 0.05$), with an effect size Cohen's $d = 0.289$ (95 % CI [0.186, 0.39]), indicating the difference between the two models reached a small effect level (Gan et al., 2025; Hong et al., 2025; Liu et al., 2026; Luo et al., 2024).

To evaluate model uncertainty, we employed resampling methods for model assessment, where resampling methods construct numerous new samples by repeatedly extracting, splitting, and disrupting existing samples with or without put-back. Due to the strong temporal correlation of the surface settlement series, MBB (Moving Block Bootstrap), a

Table 7
Statistical Test Results of Benchmark and Comparison Models.

baseline model	Comparison model	P-value	95 % CI	Cohen's d
VSA-BiLSTM (1)	BiLSTM (2)	0.0312	0.284–0.492	0.391
	SVR (3)	0.0303	0.297–0.53	0.416
	RNN (4)	0.0269	0.337–0.761	0.546
	BPNN (5)	0.0213	0.38–0.991	0.688

resampling method suitable for time series, was used to assess model performance uncertainty. The MBB method performs resampling with put-back on consecutive, overlapping data blocks from the original dataset to preserve local dependency structures, thereby enabling the estimation of statistic distributions or the construction of confidence intervals. Based on the ACF (Autocorrelation Function), the resampling block length was determined to be 15 days. Taking the land settlement data from monitoring point ML481 as an example, after 1000 resampling iterations, Model 1 yielded a mean RMSE = 1.533 mm with 95 % CI [1.32, 1.81]. Its RMSE distribution does not satisfy the normality assumption (Shapiro-Wilk $p < 0.05$), but the percentile confidence intervals do not need to depend on the distributional assumptions, which still demonstrated the model's stable predictive performance.

To fully evaluate the predictive performance and uncertainty of Model 1 at different monitoring points, we calculated the 95 % CIs for each of the five key evaluation indicators (MAE, MAPE, R^2 , RPD, and RMSE) in the same manner for each monitoring point, and added error bars to the radar plots in Fig. 13 to demonstrate the 95 % CI for each evaluation indicator. The interval shown by the error bar represents the 95 % confidence interval of the evaluation indicators, with half of the interval width representing half the width of the confidence interval, intuitively demonstrating the uncertainty of the model's performance. As shown in fig. 13, the error bar ranges for each evaluation indicator are relatively small, with a range of 0.129–0.265. Indicating that model 1 has high prediction accuracy, while having low uncertainty and strong robustness under different data divisions.

After completing model validation and statistical analysis, we further analyzed the contribution of each module (VMD, SSA, ATT) to the model's predictive performance. Using the prediction results of the VSA-BiLSTM model at monitoring point ML481 as the baseline, the changes in the model's prediction results are analyzed after respectively removing the aforementioned modules. Fig. 14 visually evaluates the model's performance after each module ablation using MAE and R^2 indicators. As shown in the figure, when the VMD, SSA, and ATT modules were individually ablated, the model's MAE increased by 0.161, 0.253, and 0.077 respectively, which indicates that the MAE value of the BiLSTM model decreases when all three modules are added individually, respectively. According to Eq. (21), the contribution percentages of VMD, SSA, and ATT to reducing the MAE of the BiLSTM model were 14.5 %, 22.81 %, and 6.94 % respectively. It is shown that all three modules contribute to the improvement of model prediction performance, with VMD and SSA contributing more, verifying the necessity of these two modules for dealing with non-stationary data and optimizing hyperparameters. After sequentially ablating the VMD, SSA, and ATT modules, the model's R^2 decreases from 0.95 to 0.93, 0.91, and 0.94 respectively, which reduces by 0.2, 0.4, and 0.1. Further evidence

demonstrates that all three modules contribute to enhancing the model's predictive performance, with VMD and SSA making greater contributions while ATT contributes to a slightly smaller extent. This was because the bidirectional structure of BiLSTM inherently possessed a certain ability to filter key time steps, making the additional gain from ATT relatively limited (6.94 %).

When the VSA (VMD-SSA-ATT) module was ablated, the MAE of the BiLSTM model increased by 0.559. Separate ablation of the VMD, SSA, and ATT modules increases the model's MAE by 0.161, 0.253, and 0.077 respectively, resulting in a cumulative error increase of 0.491, which is less than the 0.559 cumulative error when all three modules are ablated simultaneously. It suggests that simultaneous ablation of the VSA has a greater impact on model prediction accuracy, and also suggests a potential synergistic relationship between the VMD, SSA, and ATT modules.

$$\text{Percentage contribution} = \left(\frac{MAE_{\text{without module}} - MAE_{\text{full model}}}{MAE_{\text{full model}}} \right) \times 100\% \quad (21)$$

5.3. Feature visualization analysis

Each feature variable in machine learning may affect the prediction results of the model, and the degree of influence usually varies from feature to feature. In this paper, the SHAP (Shapley Additive explanation) value is used as an indicator to evaluate the relative contribution of 15 feature variables to the prediction results of the VSA-BiLSTM model (Shan et al., 2024b). In Fig. 15, the left plot shows the SHAP summary plots of the 15 feature variables, where the farther the scatter is from the baseline, the more the feature variable contributes to the model's prediction results, i.e., the higher the impact, where red and blue colors represent positive and negative contributions respectively. In the SHAP summary plot, the features are arranged along the vertical axis from top to bottom based on their impact on the model, the features at the top have a greater overall influence on the model's prediction results, while the features at the bottom have a smaller influence. The right plot is the feature importance ranking diagram, its horizontal coordinate represents the mean absolute SHAP value indicating the importance of the feature, that is, the larger the mean absolute SHAP value, the more important the feature.

As can be seen in Fig. 15, in this research project, the feature variables with the greatest influence on the model surface settlement prediction results are, in order, X7 (average tunnel depth), X8 (tunnel diameter), X11 (grouting volume), and X15 (total shield thrust). The SHAP summary plot shows that a decrease in the values of feature variables X7 and X11 has a positive contribution to the predicted surface settlement value (i.e., the surface settlement value increases), while an increase in the values of X8 and X15 has a positive contribution to the predicted settlement value. Linking to engineering practice, shallower average tunnel depth (decreased X7) intensifies soil stress redistribution, while insufficient grouting volume (decreased X11) leads to increased ground loss, both of which increase surface settlement (Cao et al., 2020; Mu et al., 2021). In addition, larger tunnel diameter (increased X8) and higher total shield thrust (increased X15) will further extend the disturbance of the ground, thus triggering greater surface settlement (Wang et al., 2019; Xie et al., 2016). The feature variables with moderate influence on the model's settlement prediction results are, in order, X12 (grouting pressure), X14 (maximum digging speed), X1, X2 (tunnel roof soil deformation modulus, cohesion), X4, X5 (tunnel floor soil deformation modulus, cohesion), X10 (cutterhead torque), and X9 (cutterhead rotation speed). The feature variables that have the least influence on the model settlement prediction results are X13 (shield machine pitch angle), X3, and X6 (Poisson's ratio of the top and bottom tunnel soils).

Although X7 (Average tunnel depth) has the highest total mean absolute SHAP value, indicating that it has the greatest influence on the

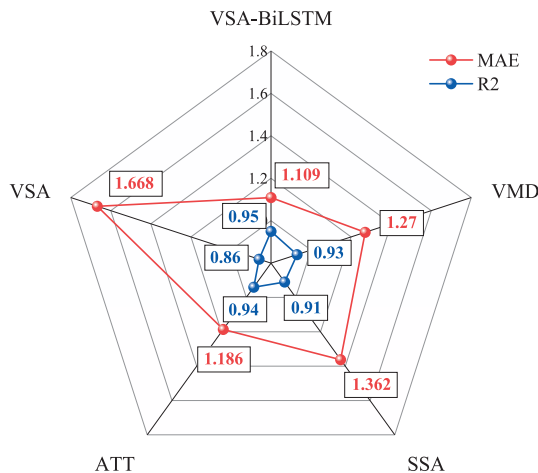


Fig. 14. VSA-BiLSTM ablation results.

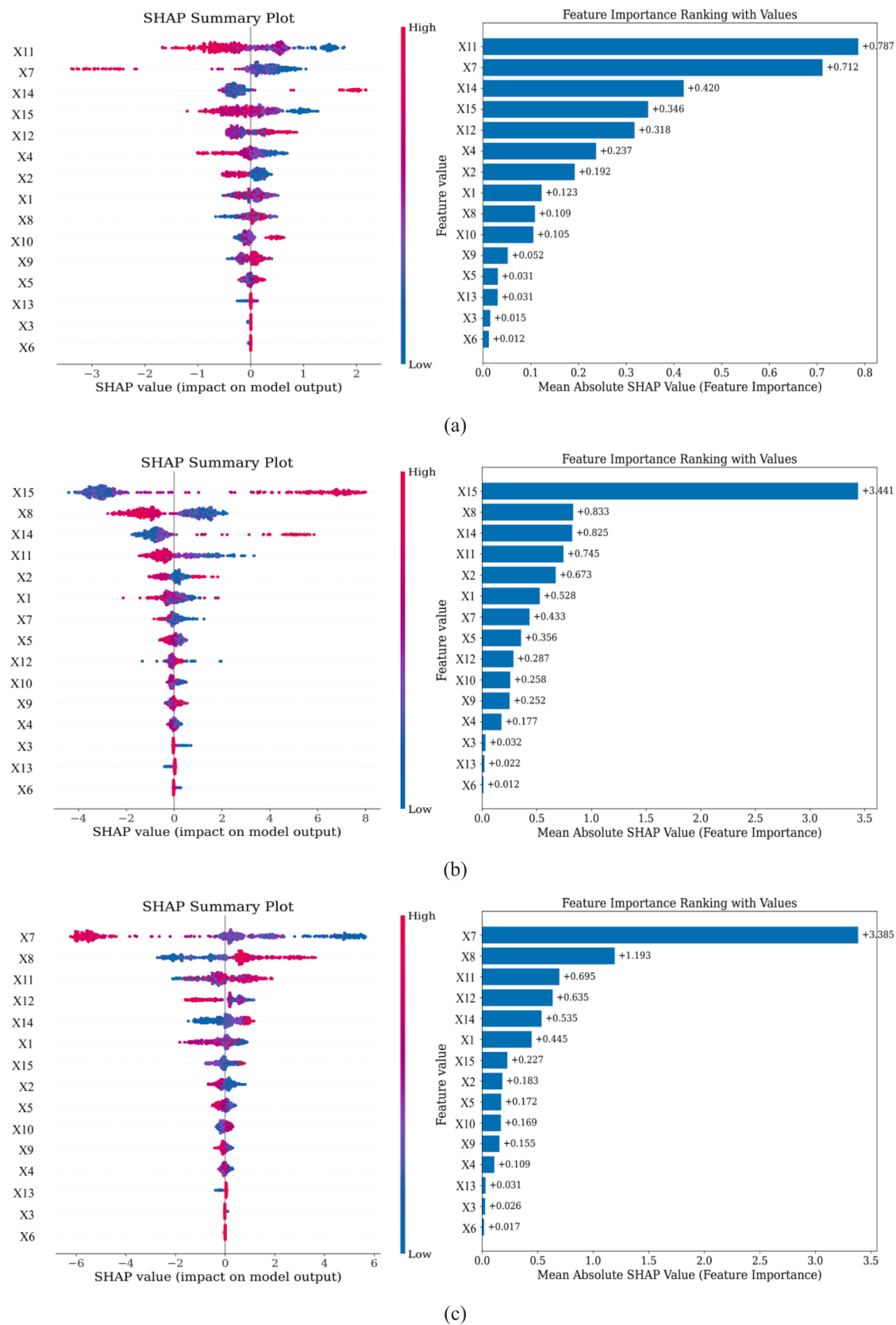


Fig. 15. SHAP values and average contribution degrees of each feature variable at three monitoring points: (a)ML29 monitoring point. (b)ML481 monitoring point. (c)ML849 monitoring point.

surface settlement prediction results. however, X7 does not always rank first in mean absolute SHAP values across the three monitoring points (ranking 2nd at ML29, 7th at ML481, and 1st at ML849), other feature variables exhibit similar characteristics, which indicates that the influence of feature variables on surface settlement is significantly affected by environmental conditions.

6. Limitations and future research

In our study, the VSA-BiLSTM model was used as the main research tool for surface settlement prediction, and satisfactory accuracy was achieved. However, our research also has some limitations which can be summarized as follows:

- (1) The four comparison models (BiLSTM, SVR, RNN, BPNN) follow the original hyperparameters, without undergoing the same optimization algorithm as model 1. Although the contribution of each module of VSA-BiLSTM to improve the accuracy of surface settlement prediction can be revealed more clearly, it may lead to the fact that the difference in prediction performance between the comparison models and the VSA-BiLSTM model partly originates from hyperparameters rather than structural differences.
- (2) The model proposed in this paper is only verified under a single working condition in Shenzhen, and its universality needs to be verified under multi-region and multi-geological condition data. Given the excellent performance of this model in this condition, the model is still expected to maintain high accuracy in other similar scenarios.
- (3) VMD defaults to the assumption that the high-frequency IMF represent noise, may determine high-frequency signals associated with mutations as noise and reject them in sudden settling scenarios. Therefore, this method is mainly suitable for slow or continuous settlement monitoring, and still needs to be verified by combining more field data and physical-mechanical models for transient and large settlement cases.

Future research could explore methods to optimize VMD, such as improving the modal decomposition strategy or introducing new noise identification mechanisms, to enable VMD to more accurately distinguish between valid high-frequency signals and noise in sudden settlement. Future research also could focus on developing more advanced interpretability tools, to address the “black box” problem of hybrid prediction models, thereby enhancing the transparency and credibility of predictive models in engineering applications. Table 8 presents the comparative analysis between this study and recent state-of-the-art research achievements in the field of surface settlement prediction.

7. Conclusion

Aiming at the surface settlement caused by shield tunneling in Shenzhen, this paper proposed a combined prediction model based on variational mode decomposition and machine learning, namely the VSA-BiLSTM model. The shield machine parameters, geological parameters, geometric parameters, 412 days of actual surface settlement data at three monitoring points, and four existing models were selected to validate the VSA-BiLSTM model:

- (1) VMD decomposed the original surface settlement sequence into multiple relatively smooth IMFs, which effectively reduced the data's non-stationarity and random noise interference. Based on the VMD-decomposed subsequences, the SSA-BiLSTM-ATT hybrid model could more fully explore the temporal feature information in each modal component, achieving better prediction results.
- (2) The SD value of the VSA-BiLSTM model was 6.39 %–30.66 % lower than other models, the R value was increased by 6.67

Table 8

Comparative analysis with state-of-the-art research.

Study	Applicable conditions	Prediction methods	Prediction results	Model advantages
(present study)	Prediction of ground settlement caused by urban subway tunnel construction	VMD + SSA + BiLSTM-ATT	mean RMSE = 1.117 mm, meanR ² = 0.95	Handling noisy data and high-accuracy prediction
(Chen et al., 2024)	Prediction of ground deformation in mining and urban areas	ARIMA + CNN-LSTM + WOA	RMSE < 3 mm, mean accuracy improvement of 28.38 % ~80.59 %	Efficient handling of complex deformation features
(Zhang et al., 2024)	Prediction of ground settlement due to coal mining	CatBoost + Intelligent optimization algorithms	SSA-CatBoost: R ² = 0.965, RMSE = 0.377, MAE = 0.297	High-accuracy prediction and feature analysis
(Ding et al., 2021)	Monitoring and Prediction of Urban Ground Settlement	MT-InSAR + Geo-Detector + LSTM	RMSE < 2.6 mm, R ² > 0.8	Long-term monitoring and trend prediction

%–31.91 %, the CRMSD was reduced by 28 %–64.7 % compared to other models, and the model still maintained a low RMSE and a high R² after 1000 resampling. It indicated that the VSA-BiLSTM model possessed higher prediction accuracy and stronger robustness than other models.

- (3) Compared with the other four models, the VSA-BiLSTM model reduced the mean values of MAE, RMSE, and MAPE by 36.1 %–58.1 %, 32.7 %–55.6 %, and 36.9 %–61.2 % respectively, while increasing the mean values of R² and RPD by 10.5 %–50.8 % and 22.2 %–130.2 % respectively, indicating that the VSA-BiLSTM model possessed higher prediction accuracy than the other models. The RMSE of VSA-BiLSTM was 32.7 % lower than that of independent BiLSTM, which proved that VMD noise reduction and SSA parameter optimization played a key role in improving the model prediction accuracy.
- (4) The ablation experiments found that the VMD, SSA, and ATT modules all improved the prediction accuracy of the model, with the VMD and SSA modules contributing more, as they reduced the MAE of the model by 14.5 % and 22.81 % respectively, and the introduction of the ATT module reduced the MAE of the model by 6.94 %. In addition, the reduction in prediction accuracy after simultaneous ablation of all three modules was higher than the cumulative reduction after ablating each module individually, suggesting a potential synergistic relationship between the three modules.
- (5) Using SHAP feature importance analysis, it was found that the average tunnel depth, total shield thrust, grouting volume and tunnel diameter had the greatest influence on the model prediction results. Grouting pressure, maximum digging speed, deformation modulus and cohesion of tunnel roof soil, deformation modulus and cohesion of tunnel floor soil, cutterhead torque, and cutterhead rotation speed had a moderate effect on the model prediction results. The shield pitch angle, the Poisson's ratio of the top soil layer of the tunnel, and the Poisson's ratio of the bottom soil layer had the least influence on the model prediction results.
- (6) The predictive accuracy of the VSA-BiLSTM model on historical data indicated its potential for early warning, but further testing was needed for real-time applications. This framework could be applied to other tunnel scenarios that required high-precision settlement prediction under noise monitoring conditions (such as subway systems and underwater tunnels).

CRedit authorship contribution statement

Na He: Writing – original draft. **Pengfei Li:** Writing – review & editing, Data curation. **Mingju Zhang:** Methodology, Conceptualization. **Chenhe Ge:** Software, Data curation.

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.

Appendix A**BiLSTM neural network**

Long Short-Term Memory (LSTM) neural networks can be used to process and predict time series data that possess long-term dependencies. Each LSTM cell is composed of 1 memory cell and 3 gating cells (forget gate, input gate, and output gate). The memory cell serves as an information conveyor belt to maintain long-term information, while the gating mechanism selectively retains critical historical information and discards unimportant information, with their combined operation enabling effective capture of long-term dependencies in time series data.

Memory cell: the core part of the LSTM unit that is responsible for saving information and transferring this information between time steps.

Forget Gate: determines what information should be discarded from a memory cell. The formula is as follows:

$$f_t = \sigma(w_{xf} \cdot x_t + w_{hf} \cdot h_{t-1} + b_f) \quad (1)$$

where f_t is the output of the oblivious gate; σ is the sigmoid function. w_{xf} and w_{hf} are the weights of the oblivious gate; x_t and h_{t-1} are the inputs at the current moment and the hidden state at the previous moment; b_f is the bias parameter of the oblivious gate.

Input gate: controls which new information should be added to the memory cell. The formula is as follows:

$$i_t = \sigma(W_{xi} \bullet x_t + W_{hi} \bullet h_{t-1} + b_i) \quad (2)$$

$$\tilde{C}_t = \tanh(W_{xc} \cdot x_t + W_{hc} \cdot h_{t-1} + b_c) \quad (3)$$

Then, by multiplying memory cells C_{t-1} and f_t from the previous moment, selectively forgetting information, and then multiplying i_t and $\tilde{C}_t$, adding new information, we obtain memory cell C_t after updating the memory state.

$$C_t = C_{t-1} \cdot f_t + \tilde{C}_t \cdot i_t \quad (4)$$

Output Gate: determines which information stored in the memory cell will become the current time step's output. The formula is as follows:

$$O_t = \sigma(W_{xo} \cdot x_t + W_{ho} \cdot h_{t-1} + b_o) \quad (5)$$

$$h_t = O_t \cdot \tanh(C_t) \quad (6)$$

Bidirectional Long Short-Term Memory (BiLSTM) consists of two LSTM layers with opposing information flow directions, enabling the model to simultaneously process temporal data in both forward (input-to-output) and backward (output-to-input) directions. Through this architecture, BiLSTM can obtain more comprehensive contextual information in the whole time series, which overcomes the limitation of LSTM networks in insufficient feature extraction from data.

Appendix B**Table 9**

The center frequency of the intrinsic mode function.

K	IMF1	IMF2	IMF3	IMF4	IMF5	IMF6	IMF7
2	0.007	0.135					
3	0.006	0.097	0.217				
4	0.005	0.054	0.155	0.233			
5	0.004	0.042	0.139	0.181	0.274		
6	0.004	0.033	0.134	0.16	0.234	0.319	
7	0.004	0.029	0.126	0.147	0.206	0.277	0.324

Appendix C**VMD-SSA Module Pseudocode**

Part 1: VMD Denoising

Input: Original land settlement sequence

Output: Denoised settlement sequence

Set VMD parameters: K, α , tol, tau

VMD decomposition:

Obtain K IMF components set

For each IMF component:

A. Calculate wk

(continued on next page)

(continued)

VMD-SSA Module Pseudocode

B. Calculate correlation coefficient

4. Reconstruct denoised signal:

If determined to be noise, discard that IMF component

5. Output: Denoised settlement sequence

Part 2: SSA optimization

Input: Denoised settlement sequence6. Define parameters: N, T_max, D_ratio, O_ratio, V_ratio

7. Initialize population:

Initialize BiLSTM hyperparameters (epochs, batch_size, dropout, hidden_layers, hidden_units)

8. For t from 1 to T_max:

a. Sort the population by fitness

b. Update positions

Output: the optimal hyperparameter combination

Data availability

Data will be made available on request.

References

- Cao, L., Zhang, D., Fang, Q., Yu, L., 2020. Movements of ground and existing structures induced by slurry pressure-balance tunnel boring machine (SPB TBM) tunnelling in clay. *Tunn. Undergr. Space Technol.* 97, 103278.
- Chen, G., Tang, B., Zeng, X., Zhou, P., Kang, P., Long, H., 2022. Short-term wind speed forecasting based on long short-term memory and improved BP neural network. *Int. J. Electr. Power Energy Syst.* 134, 107365.
- Chen, Y., Chen, X., Guo, S., Li, H., Du, P., 2024. A novel surface deformation prediction method based on AWC-LSTM model. *Int. J. Appl. Earth Obs. Geoinf.* 135, 104292.
- Ding, Q., Shao, Z., Huang, X., Altan, O., Zhuang, Q., Hu, B., 2021. Monitoring, analyzing and predicting urban surface subsidence: a case study of wuhan city, China. *Int. J. Appl. Earth Obs. Geoinf.* 102, 102422.
- Dragomiretskiy, K., Zosso, D., 2014. Variational mode decomposition. *IEEE Trans. Signal Process.* 62, 531–544.
- Duan, X., 2025. Settlement prediction of nanjing metro line 10 with HOA-VMD-LSTM. *Measurement* 244, 116477.
- Fu, W., Wang, K., Li, C., Tan, J., 2019. Multi-step short-term wind speed forecasting approach based on multi-scale dominant ingredient chaotic analysis, improved hybrid GWO-SCA optimization and ELM. *Energy. Conver. Manage.* 187, 356–377.
- Gan, W., Ma, R., Zhao, W., Peng, X., Cui, H., Yan, J., Duan, S., Wang, L., Feng, P., Chu, J., 2025. A VMD-LSTNet-attention model for concentration prediction of mixed gases. *Sens. Actuators B* 422, 136641.
- Gan, W., 2025. Predicting the surface temperature of radiative cooling coatings with time-series forecasting: validation of the small-batch training dataset and implementation of a hyperparameter optimization strategy.
- Gao, X., Guo, W., Mei, C., Sha, J., Guo, Y., Sun, H., 2023. Short-term wind power forecasting based on SSA-VMD-LSTM. *Energy Rep.* 9, 335–344.
- Ge, C., Li, P., Zhang, M., Yang, M., 2025. Identification of surface subsidence risk in deep foundation pits using a mamba fusion model. *Eng. Appl. Artif. Intel.* 161, 112077.
- Gong, T., Yuan, X., Yuan, Y., Lei, X., Wang, X., 2019. Application of tentative variational mode decomposition in fault feature detection of rolling element bearing. *Measurement* 135, 481–492.
- Hong, J.-T., Han, S., Yan, J., Liu, Y.-Q., 2025. Dual-path frequency mamba-transformer model for wind power forecasting. *Energy* 332, 137225.
- Huang, H., Chen, J., Huo, X., Qiao, Y., Ma, L., 2021. Effect of multi-scale decomposition on performance of neural networks in short-term traffic flow prediction. *IEEE Access* 9, 50994–51004.
- Huang, L., Xin, J., Jiang, Y., Tang, Q., Zhang, H., Yang, S.X., Zhou, J., 2025. Bridge temperature data extraction and recovery based on physics-aided VMD and temporal convolutional network. *Eng. Struct.* 331, 119967.
- Ju, H., Shi, H., Shen, W., Deng, Y., 2024. An accurate and low-cost vehicle-induced deflection prediction framework for long-span bridges using deep learning and monitoring data. *Eng. Struct.* 310, 118094.
- Kang, Q., Chen, E.J., Li, Z.-C., Luo, H.-B., Liu, Y., 2023. Attention-based LSTM predictive model for the attitude and position of shield machine in tunneling. *Underground Space* 13, 335–350.
- Lei, L., Shao, S., Liang, L., 2024. An evolutionary deep learning model based on EWKM, random forest algorithm, SSA and BiLSTM for building energy consumption prediction. *Energy* 288, 129795.
- Lei, Y., Lin, J., He, Z., Zuo, M.J., 2013. A review on empirical mode decomposition in fault diagnosis of rotating machinery. *Mech. Syst. Sig. Process.* 35, 108–126.
- Li, C., Liu, Y., Liao, Y., Wang, J., 2022. A VME method based on the convergent tendency of VMD and its application in multi-fault diagnosis of rolling bearings. *Measurement* 198, 111360.
- Lin, J., Qu, L., 2000. Feature extraction based on morlet wavelet and its application for mechanical fault diagnosis. *J. Sound Vib.* 234, 135–148.
- Liu, F., Liang, C., 2024. Short-term power load forecasting based on AC-BiLSTM model. *Energy Rep.* 11, 1570–1579.
- Liu, L., Zhou, W., Gutierrez, M., 2022. Effectiveness of predicting tunneling-induced ground settlements using machine learning methods with small datasets. *J. Rock Mech. Geotech. Eng.* 14, 1028–1041.
- Liu, X., Li, K., Jiang, A., Fang, Q., Zhang, R., 2024. Prediction interaction responses between railway subgrade and shield tunnelling using machine learning with sparrow search algorithm. *Transp. Geotech.* 44, 101169.
- Liu, Y., Yang, C., Huang, K., Gui, W., 2020. Non-ferrous metals price forecasting based on variational mode decomposition and LSTM network. *Knowl.-Based Syst.* 188, 105006.
- Liu, Z., Liu, H., 2023. A novel hybrid model based on GA-VMD, sample entropy reconstruction and BiLSTM for wind speed prediction. *Measurement* 222, 113643.
- Liu, Z., Liu, Y., Zhang, Y., Wu, C., Zhang, S., Sun, C., 2026. Data-driven lithium-ion battery SOH prediction: a novel SHMM-transformer-BiGRU hybrid neural network method. *Measurement* 257, 118579.
- Luo, Q., He, S., Han, X., Wang, Y., Li, H., 2024. LSTTN: a long-short term transformer-based spatiotemporal neural network for traffic flow forecasting. *Knowl.-Based Syst.* 293, 111637.
- Lu, P., Ye, L., Pei, M., Zhao, Y., Dai, B., Li, Z., 2022. Short-term wind power forecasting based on meteorological feature extraction and optimization strategy. *Renew. Energy* 184, 642–661.
- Mu, B., Xie, X., Li, X., Li, J., Shao, C., Zhao, J., 2021. Monitoring, modelling and prediction of segmental lining deformation and ground settlement of an EPB tunnel in different soils. *Tunn. Undergr. Space Technol.* 113, 103870.
- Nie, Y., Zhang, Q., Du, Y., Du, L., 2025. TBM jamming risk prediction method based on fuzzy theory and bi-LSTM. *Tunn. Undergr. Space Technol.* 157, 106329.
- Qin, C., Huang, G., Yu, H., Zhang, Z., Tao, J., Liu, C., 2024. Adaptive VMD and multi-stage stabilized transformer-based long-distance forecasting for multiple shield machine tunneling parameters. *Autom. Constr.* 165, 105563.
- Shan, F., He, X., Armaghani, D.J., Sheng, D., 2024a. Effects of data smoothing and recurrent neural network (RNN) algorithms for real-time forecasting of tunnel boring machine (TBM) performance. *J. Rock Mech. Geotech. Eng.* 16, 1538–1551.
- Shan, F., He, X., Jahed Armaghani, D., Xu, H., Liu, X., Sheng, D., 2024b. Real-time forecasting of TBM cutterhead torque and thrust force using aware-context recurrent neural networks. *Tunn. Undergr. Space Technol.* 152, 105906.
- Shen, L., Zhou, P., Han, Y., Mi, L., Xu, W., Xu, J., 2025. Wind farm layout optimization based on dynamic levy sparrow search algorithm: a multi-parameter analysis with active yaw control. *Energy* 332, 135989.
- Song, D., Dai, F., Liu, Y., Tan, H., Wei, M., 2025. An efficiently designed CNN-transformer fusion network for automatic and real-time microseismic signal classification. *Tunn. Undergr. Space Technol.* 161, 106534.
- Su, T., Liu, D., Cui, X., Dou, X., Lei, B., Cheng, X., Yuan, M., Chen, R., 2024. Prediction of DEDI index for meteorological drought with the VMD-CBiLSTM hybrid model. *J. Hydrol.* 641, 131805.
- Tian, Q., Cao, M., Chen, S., Yin, H., 2021. Structure-exploiting discriminative ordinal multioutput regression. *IEEE Trans. Neural Netw. Learning Syst.* 32, 266–280.
- Wang, G., Fang, Q., Du, J., Wang, J., Li, Q., 2023a. Deep learning-based prediction of steady surface settlement due to shield tunnelling. *Autom. Constr.* 154, 105006.
- Wang, G., Fang, Q., Wang, J., Li, Q., Song, H., Huang, J., 2025. Artificial intelligence prediction of surface settlement induced by twin shields tunnelling. *Tunn. Undergr. Space Technol.* 161, 106606.
- Wang, J., Fang, Q., Zheng, G., Wang, G., Chen, J., Zhang, J., Zhao, P., 2025. Influences of bit button wear on performance of rotary-percussive drilling: MBD-DEM coupling simulation and verification. *J. Rock Mech. Geotech. Eng.* 17, 1585–1598.
- Wang, L., Geng, P., Chen, J., Wang, T., 2023b. Machine learning-based fragility analysis of tunnel structure under different impulsive seismic actions. *Tunn. Undergr. Space Technol.* 133, 104953.
- Wang, R., Li, C., Fu, W., Tang, G., 2020. Deep learning method based on gated recurrent unit and variational mode decomposition for short-term wind power interval prediction. *IEEE Trans. Neural Netw. Learning Syst.* 31, 3814–3827.
- Wang, Y., Markert, R., Xiang, J., Zheng, W., 2015. Research on variational mode decomposition and its application in detecting rub-impact fault of the rotor system. *Mech. Syst. Sig. Process.* 60–61, 243–251.
- Wen, Y., Pan, S., Li, X., Li, Z., 2024. Highly fluctuating short-term load forecasting based on improved secondary decomposition and optimized VMD. *Sustainable Energy Grids Networks* 37, 101270.

- Wang, Z., Yao, W., Cai, Y., Xu, B., Fu, Y., Wei, G., 2019. Analysis of ground surface settlement induced by the construction of a large-diameter shallow-buried twin-tunnel in soft ground. *Tunn. Undergr. Space Technol.* 83, 520–532.
- Xie, Q., Ma, L., Liu, Y., Fu, Q., Shen, Z., Wang, X., 2023. An improved SSA-BiLSTM-based short-term irradiance prediction model via sky images feature extraction. *Renew. Energy* 219, 119507.
- Xie, X., Yang, Y., Ji, M., 2016. Analysis of ground surface settlement induced by the construction of a large-diameter shield-driven tunnel in shanghai, China. *Tunn. Undergr. Space Technol.* 51, 120–132.
- Xu, D., Wang, Y., Huang, J., Liu, S., Xu, S., Zhou, K., 2023a. Prediction of geology condition for slurry pressure balanced shield tunnel with super-large diameter by machine learning algorithms. *Tunn. Undergr. Space Technol.* 131, 104852.
- Xu, Q., Huang, X., Zhang, B., Zhang, Z., Wang, J., Wang, S., 2023b. TBM performance prediction using LSTM-based hybrid neural network model: Case study of baimang river tunnel project in shenzhen, China. *Underground Space* 11, 130–152.
- Xu, Z., Che, A., Zhou, H., Shen, Y., He, W., 2024. Optimized deep learning modelling for predicting the diffusion range and state change of filling projects. *Tunn. Undergr. Space Technol.* 154, 106073.
- Ye, X.-W., Ma, S.-Y., Liu, Z.-X., Chen, Y.-B., Lu, C.-R., Song, Y.-J., Li, X.-J., Zhao, L.-A., 2024a. LSTM-based deformation forecasting for additional stress estimation of existing tunnel structure induced by adjacent shield tunneling. *Tunn. Undergr. Space Technol.* 146, 105664.
- Ye, X.-W., Wei, Y.-J., Zhang, X.-L., Ding, Y., Chen, Y.-M., 2024b. Confining pressure forecasting of shield tunnel lining during construction based on LSTM-PSO models combined with the multi-output recursive strategy. *Tunn. Undergr. Space Technol.* 143, 105499.
- Yu, S., Zhang, Z., Wang, S., Huang, X., Lei, Q., 2024. A performance-based hybrid deep learning model for predicting TBM advance rate using attention-ResNet-LSTM. *J. Rock Mech. Geotech. Eng.* 16, 65–80.
- Zhang, B., Xu, C., Dai, X., Xiong, X., 2024a. Research on mining land subsidence by intelligent hybrid model based on gradient boosting with categorical features support algorithm. *J. Environ. Manage.* 354, 120309.
- Zhang, J., Xin, X., Shang, Y., Wang, Y., Zhang, L., 2023. Nonstationary significant wave height forecasting with a hybrid VMD-CNN model. *Ocean Eng.* 285, 115338.
- Zhang, Y., Ren, X., Zhang, J., Zhang, Y., Ma, Z., 2024b. A novel workflow including denoising and hybrid deep learning model for shield tunneling construction parameter prediction. *Eng. Appl. Artif. Intel.* 133, 108103.
- Zhao, L., Li, Z., Qu, L., Zhang, J., Teng, B., 2023. A hybrid VMD-LSTM/GRU model to predict non-stationary and irregular waves on the east coast of China. *Ocean Eng.* 276, 114136.
- Zhou, D., Zhuang, X., Zuo, H., Cai, J., Zhao, X., Xiang, J., 2022. A model fusion strategy for identifying aircraft risk using CNN and att-BiLSTM. *Reliab. Eng. Syst. Saf.* 228, 108750.
- Zhou, X., Hu, W., Zhang, Z., Ye, J., Zhao, C., Bian, X., 2024. Adaptive mutation sparrow search algorithm-elman-AdaBoost model for predicting the deformation of subway tunnels. *Underground Space* 17, 320–360.
- Zou, X., Zeng, J., Yan, G., Mohammed, K.J., Abbas, M., Abdullah, N., Elattar, S., Khadimallah, M.A., Toghrol, S., Escorcio-Gutierrez, J., 2024. Advancing tunnel equipment maintenance through data-driven predictive strategies in underground infrastructure. *Comput. Geotech.* 173, 106532.