IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 26 February 2025, accepted 12 March 2025, date of publication 17 March 2025, date of current version 28 March 2025.

Digital Object Identifier 10.1109/ACCESS.2025.3551730

RESEARCH ARTICLE

Landslide Displacement Prediction Model Based on Optimal Decomposition and Deep Attention Mechanism

SHUAI REN[1,2,3], KAMARUL HAWARI GHAZALI[1], (Senior Member, IEEE), YUANFA JI[1,2,3], AND SAMRA UROOJ KHAN[1]

[1]Faculty of Electrical and Electronics Engineering Technology, Universiti Malaysia Pahang Al-Sultan Abdullah, Pekan, Pahang 26600, Malaysia
[2]Information and Communication School, Guilin University of Electronic Technology, Guilin 541004, China
[3]Guangxi Key Laboratory of Precision Navigation Technology and Application, Guilin University of Electronic Technology, Guilin 541004, China

Corresponding authors: Kamarul Hawari Ghazali (kamarul@ump.edu.my) and Yuanfa Ji (jiyuanfa@163.com)

This work was supported in part by Guangxi Science and Technology Plan Project under Grant Guike AA230620388, Grant Guike AA24206043, Grant Guike AB23026120, and Grant Guike ZY23055048; in part by the National Natural Science Foundation of China under Grant U23A20280, Grant 62161007, and Grant 62471153; in part by Nanning Science Research and Technology Development Plan under Grant 20231029 and Grant 20231011; in part by Guangxi Autonomous Region Major Talent Project; in part by the Collaborative Innovation Center for Beidou Positioning Services and Border and Coastal Defense Security Applications; and in part by the Graduate Innovation Project of Guilin University of Electronic Technology.

ABSTRACT Landslide displacement forecasting is crucial for disaster prevention and risk management, as it enables timely warnings and effective mitigation strategies. However, the highly nonlinear and complex nature of landslide displacement poses significant challenges for accurate prediction. To address this, this study proposes an advanced forecasting framework integrating the Chebyshev Levy Flight-Sparrow Search Algorithm (CLF-SSA) with Variational Mode Decomposition (VMD) to enhance decomposition accuracy and optimize parameter selection. The trend component is modeled using the Autoregressive Integrated Moving Average (ARIMA) with a grid search strategy, while the periodic component is predicted using a Bidirectional Long Short-Term Memory network with an Attention mechanism (BiLSTM-Attention), which dynamically adjusts the contribution of influencing factors. Grey Relational Analysis (GRA) is further employed to identify key external driving factors, enhancing prediction accuracy. Experimental results demonstrate that the proposed model significantly improves predictive performance, reducing the Root Mean Square Error (RMSE) by 60% compared to the traditional XGBoost model and by 33% compared to the Empirical Mode Decomposition-BiLSTM (EMD-BiLSTM) model. Moreover, the Mean Absolute Scaled Error (MASE) analysis confirms the robustness of the model in capturing both short-term fluctuations and long-term trends. Given its superior predictive accuracy and practical applicability, this approach provides valuable technical support for landslide monitoring and early warning systems.

INDEX TERMS Landslide displacement prediction, intelligent optimization algorithm, variational mode decomposition, bidirectional long short term memory network, attention mechanism.

I. INTRODUCTION

A landslide refers to the sliding phenomenon of the rock and soil mass on the slope along a specific shear surface under the influence of internal and external active factors. Its occurrence is marked by high frequency and great harmfulness [1], posing a serious threat to the life and property safety of people around the world and significantly impeding the social development process. China is among the countries most severely affected by landslide disasters globally. For instance, on September 10, 2023, heavy rain in multiple areas of Yulin, Guangxi, trigg-ered landslides, resulting in 7 fatalities and 3 people missing. The direct economic loss amounted to as high as 28.5294 million yuan. According to the statistics of the National Bureau of Statistics, there were as many as 925 landslide disasters in 2023 [2], causing

The associate editor coordinating the review of this manuscript and approving it for publication was Jiajie Fan.

severe damage to a large number of lives and properties. Constructing a landslide displacement prediction model is the key and core technology for realizing the monitoring and early warning of landslide geological disasters and can effectively mitigate the significant losses caused by disasters [3]. Therefore, conducting research on high-precision landslide displacement prediction models has extremely crucial practical significance [4].

Traditional landslide displacement prediction techniques contain empirical analysis, statistical methods and mechanism studies [5]. With the development of nonlinear theories such as machine learning and neural networks, more and more scholars apply them to landslide displacement prediction research [6]. In order to improve the interpretability and accuracy of the prediction model, many studies have introduced the time series decomposition method, which decomposes the landslide displacement into the trend displacement determined by the geological conditions and the cycle displacement caused by the external seasonal influences [7], which helps to identify the influence of different factors on the landslide changes by predicting each decomposition sequence individually, and thus increases the rationality and accuracy of the model. Huang et al used double exponential smoothing method to decompose landslide displacements, which effectively improved the fitting effect of trend term displacements by uneven weighting [8]. However, this method struggles to capture periodic fluctuations effectively and exhibits limited adaptability to abrupt changes, potentially leading to prediction lags and increased errors. Meng et al introduced Empirical Mode Decomposition (EMD) adaptive decomposition algorithm, which flexibly adapts to the changing characteristics of the data in different time periods, and is easy to remove the high-frequency noise, so as to improve the decomposition accuracy [9]. However, EMD is prone to modal aliasing when dealing with sudden signal changes [10]. To solve this problem, Wang et al added white noise to the original sequence and proposed the Ensemble Empirical Mode Decomposition (EEMD) decomposition algorithm, which enhances the scale adaptability of the signal and makes the boundary of the Intrinsic Mode Function (IMF) component clearer [11]. However, EEMD still requires multiple realizations to average out the effects of the added noise, making it computationally expensive and sensitive to parameter selection. In pursuit of a more stable and efficient decomposition method, Li et al introduced Variational Mode Decomposition (VMD), which formulates the decomposition process as a constrained optimization problem in the frequency domain. Unlike EMD and EEMD, which rely on iterative sifting processes, VMD directly extracts modes with distinct frequency characteristics, effectively mitigating mode mixing and improving decomposition reliability. Moreover, VMD exhibits strong anti-noise capabilities, ensuring robustness in noisy environments, and achieves higher computational efficiency, making it particularly suitable for processing large-scale landslide displacement data [12]. Given the nonlinear and multi-scale nature of landslide displacement, VMD provides a more structured and theoretically grounded approach for separating trend and periodic components. Despite its advantages, the effectiveness of VMD heavily depends on parameter selection, particularly the penalty factor and the number of decomposition modes, which influence both decomposition accuracy and computational efficiency. Traditionally, these parameters are determined through empirical tuning or exhaustive experiments, leading to inefficiencies and reduced interpretability [13]. To overcome this limitation, Jiang et al integrated VMD with Particle Swarm Optimization (PSO) to automatically optimize key parameters, reducing subjective bias and improving the generalization ability of the decomposition process [14]. However, conventional metaheuristic optimization methods such as PSO may suffer from premature convergence and suboptimal parameter tuning when handling high-dimensional optimization problems.

Building upon this foundation, this study introduces an improved optimization framework that integrates VMD with the Chebyshev Levy Flight-Sparrow Search Algorithm (CLF-SSA). The CLF-SSA algorithm enhances optimization efficiency by incorporating Chebyshev chaotic mapping for population initialization, ensuring a more uniform and diverse distribution of initial solutions, which prevents premature convergence and improves global search capability. Additionally, the Levy flight strategy is embedded into the SSA framework to enhance exploration by enabling large stochastic jumps, allowing the algorithm to escape local optima and conduct a more comprehensive search of the parameter space. Furthermore, the adaptive search mechanism of SSA dynamically balances exploration and exploitation through the interaction between discoverers, followers, and vigilantes, facilitating efficient fine-tuning of VMD parame-ters and leading to more accurate decomposition results while maintaining computational efficiency. By dynamically optimizing VMD parameters, CLF-SSA mitigates the limita-tions of conventional optimization approaches, enhancing both decomposition accuracy and computational efficiency. Experimental comparisons with existing decomposition methods demonstrate that the CLF-SSA-VMD framework achieves superior predictive accuracy by effectively extracting trend and periodic components while maintaining robustness in noisy environments. This study provides a novel approach to landslide displacement decomposition and prediction, offering improved methodological robustness and practical applicability for landslide monitoring and early warning systems.

In displacement prediction, machine learning models like Random Forests [15] and Support Vector Regression (SVR) [16] have been used but struggle with the complexity of landslide systems [17]. Neural networks such as Back Propa-gation Neural Networks (BPNN) [18] and Extreme Learning Machines (ELM) [19] improve prediction accuracy but fail to capture temporal-spatial relationships [20]. Recurrent Neural Networks (RNN) address this through hidden states that model temporal dependencies [21], though

traditional RNN face gradient issues. Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) models enhance long-term prediction capabilities via gating mechanisms [22]. Advanced variants, such as LSTM-FC [23], Prophet-LSTM [24], and Bidirectional Long Short-Term Memory (BiLSTM) [25], have demonstrated superior accuracy by incorporating additional context, such as groundwater levels and rainfall. Hybrid models like Convolutional Neural Network - Bidirectional Long Short-Term Memory (CNN-BiLSTM) combine spatial feature extraction with temporal modeling to achieve better predictive performance [26].

The attention mechanism is very popular in the field of machine translation. The traditional encoder-decoder architecture will compress the entire input sequence into a fixed-length context vector, which may lead to information loss. The attention mechanism uses a dynamic calculation weight method to flexibly extract information and reduce Information compression loss [27]. The application of this type of method to the time series field has significantly improved the model's ability to handle long-distance dependencies, feature extrac-tion, model interpretability and generalization capabilities.

In summary, this paper designs an optimized decomposition CLF-SSA-VMD method to accurately extract the trend and periodic changes of landslides, introduces simple Auto-regressive Integrated Moving Average (ARIMA) machine learning to fit the landslide trend part, and improves the prediction efficiency. A BiLSTM-Attention integrated model is constructed to predict the landslide period term, and then the total displacement prediction result is calculated based on the time series addition model. The effectiveness of the model in this paper is verified by experimental comparison of three evaluation indicators and four prediction models in the Bai-shui River area.

II. THEORY AND METHODS

A. IMPROVED SPARROW SEARCH ALGORITHM BASED ON CHAOTIC MAP LEVY FLIGHT

1) SPARROW SEARCH ALGORITHM

Sparrow Search Algorithm (SSA) consists of finders and joiners. The finder is responsible for guiding the direction of the population and finding food sources, while the joiner observes the state of the finder and uses the finder's position to find food. The ratio of the two roles remains constant in the population, but the role identity is dynamic. Joiners can become finders by finding a better food location [28]. The specific steps of the algorithm are as follows:

Step 1, initialize the position of each sparrow and calculate the individual fitness value;

$$\boldsymbol{F}_x = \begin{bmatrix} f([\boldsymbol{X}_{1,1}, \boldsymbol{X}_{1,2}, \cdots, \boldsymbol{X}_{1,d}]) \\ f([\boldsymbol{X}_{2,1}, \boldsymbol{X}_{2,2}, \cdots, \boldsymbol{X}_{2,d}]) \\ \vdots \\ \vdots \\ f([\boldsymbol{X}_{n,1}, \boldsymbol{X}_{n,2}, \cdots, \boldsymbol{X}_{n,d}]) \end{bmatrix} \tag{1}$$

where f is the fitness value; n is the number of sparrows; d is the dimension of the variable to be solved;

Step 2, discoverer location update;

$$\boldsymbol{X}_{i,j}^{t+1} = \begin{cases} \boldsymbol{X}_{i,j}^{t} \cdot \exp\left(-\frac{i}{\alpha \cdot \text{iter}_{\max}}\right) & R_2 < ST \\ \boldsymbol{X}_{i,j}^{t} + \boldsymbol{Q} \cdot \boldsymbol{L} & R_2 \geq ST \end{cases} \tag{2}$$

where $\boldsymbol{X}_{i,j}^{t+1}$ represents the next position of the discoverer; j ranges from [1, d]; t represents the current iteration number. When $R_2 < ST$, it means that the population environment is very safe and the discoverer can further expand the search area. When $R_2 \geq ST$, it means that there is danger around the population and the discoverer needs to lead the population to a safe area.

Step 3, joiner location update;

$$\boldsymbol{X}_{i,j}^{t+1} = \begin{cases} Q \cdot \exp(-\frac{\boldsymbol{X}_{worst} - \boldsymbol{X}_{i,j}^{t}}{\alpha \cdot \text{iter}_{\max}}) & i > n/2 \\ \boldsymbol{X}_{p}^{t+1} + \left|\boldsymbol{X}_{i,j}^{t} - \boldsymbol{X}_{p}^{t+1}\right| \cdot \boldsymbol{A}^{+} \cdot \boldsymbol{L} & otherwise \end{cases} \tag{3}$$

where $\boldsymbol{X}_{worst}$ represents the most dangerous area selected by the population; $\boldsymbol{X}_p$ represents the safest area selected by the discoverer. When the condition is $i > n/2$, it means that the i-th joiner did not find food, so he chose to go somewhere else to seek food. Under other conditions, the joiner preys according to the location of the discoverer.

Step 4: When the population is aware of the danger, sparrows at the edge of the population quickly move toward the safe area, while sparrows in the middle of the population move randomly and converge with other sparrows to the safe area;

$$\boldsymbol{X}_{i,j}^{t+1} = \begin{cases} \boldsymbol{X}_{best}^{t} + \beta \cdot \left|\boldsymbol{X}_{i,j}^{t} - \boldsymbol{X}_{best}^{t}\right| & f_i > f_w \\ \boldsymbol{X}_{i,j}^{t} + K \cdot (\frac{\left|\boldsymbol{X}_{i,j}^{t} - \boldsymbol{X}_{worst}^{t}\right|}{(f_i - f_w) + \varepsilon}) & f_i = f_g \end{cases} \tag{4}$$

where $\boldsymbol{X}_{best}$ represents the safest position selected by the population; β represents a normally distributed random number; K represents a random number in the range of [1, -1]; ε represents a very small parameter, eliminating the meaningless situation where the denominator is 0.

Step 5, determine whether the end condition is met. If so, output the global optimal value and the best fitness. If not, return to step 2.

2) CHEBYSHEV CHAOTIC MAP INITIALIZATION POPULATION

Chaotic mapping is a nonlinear theoretical method with the characteristics of randomness, complexity and unpredictability [29]. The introduction of chaotic mapping improves population diversity, expands the global search space, and makes it easier to jump out of the local optimal solution. The sequence generated by chaos has direction and regularity, which effectively improves the efficiency of local search [30] and accelerates the convergence of the algorithm. At the same time, the complexity and unpredictability of chaotic mapping enable the optimization algorithm to maintain a good search ability under different initial conditions, thereby improving the robustness of the optimization algorithm.

The traditional SSA algorithm uses a random method to initialize the population. During the traversal process, the algorithm tends to converge prematurely and fall into a local optimal solution, resulting in poor optimization ability. To address this problem, this paper introduces the Chebyshev chaos mapping method to initialize the population. The expression is as follows:

$$X_{n+1} = \cos(b \cos^{-1} X_n) \tag{5}$$

where b is a constant. After many experiments, it is found that when Chebyshev chaotic mapping $b = 4$, the ability of SSA algorithm to jump out of local optimal solution is effectively improved. Both random method and Chebyshev mapping method are iterated 1000 times, and the results are shown in the following Fig 1:

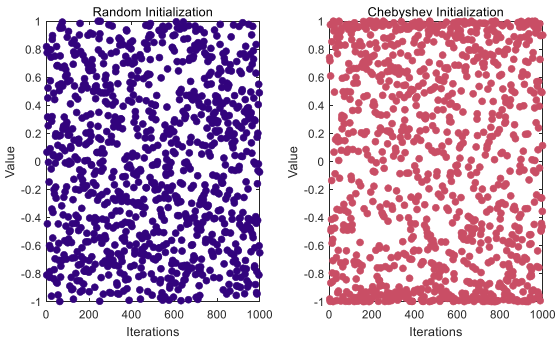


FIGURE 1. Chebyshev chaotic sequence simulation distribution.

As shown in Fig 1, Chebyshev initialization makes the population more evenly distributed and improves the convergence speed. The population size at the boundary is relatively large, which improves the global search capability.

3) LEVY FLIGHT IMPROVED POSITION UPDATE

The Levy flight strategy has the characteristics of random walk, occasionally generating a large step size in a short period of time to explore the optimal solution [31], enriching the diversity of the population and improving the ability to find the optimal solution. Chebyshev initialization improves the global search ability in the early stage of iteration, but in the middle of the iteration, due to the characteristics of the algorithm itself, the step size is small and it is easy to fall into the local optimum. To solve this problem, this paper introduces the Levy flight strategy to update the position of the joiner. The improved position formula is:

$$X_{i,j}^{t+1} = \begin{cases} Q \cdot \exp\left(\dfrac{X_{worst}(t) - X_{i,j}^{t}}{i^2}\right), i > N/2 \\ X_{best}(t) + X_{best}(t) \otimes Levy(e) * 100 \text{ other} \end{cases} \tag{6}$$

where $Levy(e)$ can be expressed as:

$$Levy(e) = 0.01 \times \frac{r_1 \times u}{|r_2|^{(1/h)}} \tag{7}$$

where r_1 and r_2 are random numbers that obey the normal distribution and range from [0, 1], h is a constant, which is 1.5 in this paper, and u can be expressed as:

$$u = \left(\frac{\Gamma(1+h) \times \sin(\pi h/2)}{\Gamma((1+h)/2) \times h \times 2^{((h-1)/2)}}\right) \tag{8}$$

where, Γ represents the gamma function.

B. POPULATION INTELLIGENCE ALGORITHM TO OPTIMIZE VMD

1) VMD

VMD is an adaptive, fully non-recursive modal decomposition method that can effectively solve the problems of modal aliasing and endpoint effects that occur during the decomposition process [32]. VMD fixes the number of modal components obtained by decomposition through parameter setting and exhibits strong adaptability when processing non-stationary data. It uses the Gaussian smoothing method combined with the squared gradient criterion to demodulate the signal and transforms the target problem into a constrained variational problem for solution.

$$\begin{aligned} &\min_{\{u_k\},\{w_k\}} \sum_k ||\partial(t)[(\delta(t) + \frac{\mathrm{j}}{\pi t}) * u_k(t)]e^{-\mathrm{j}w_k t}||^2 \\ &s.t. \sum_k u_k(t) = f(t) \end{aligned} \tag{9}$$

where $u_k(t)$ represents the modal function, $\{w_k\}$ represents the center frequency of the kth modal function, $\{u_k\}$ represents the kth modal function component;

The Lagrange multiplication operator is introduced to solve equation (9), which is expressed as equation (10):

$$\begin{aligned} L(\{u_k\}, \{w_k\}, \lambda) = a \sum_k ||\partial_t[(\delta(t) \\ + \frac{\mathrm{j}}{\pi t})u_k(t)]e^{-\mathrm{j}w_k t}||^2 + ||f(t) - \sum_k u_k(t)||^2 \\ + \left\langle \lambda(t), f(t) - \sum_k u_k(t) \right\rangle \end{aligned} \tag{10}$$

where: a is the quadratic penalty factor; $\lambda(t)$ is the Lagrange multiplier;

Finally, the updated expression of the center frequency is obtained as formula (11):

$$w_k^{N+1} = \frac{\int_0^\infty w \left|\hat{u_k}(w)\right|^2 dw}{\int_0^\infty \left|\hat{u_k}(w)\right|^2 dw} \tag{11}$$

where $\hat{u_k}(w)$ represents the Wiener filter of the current residual component, w_k^{N+1} represents the center of gravity of the power spectrum of the current modal function.

2) CLF-VMD

The decomposition effect of VMD depends on parameter selection. Inappropriate parameter settings may lead to over-decomposition or under-decomposition of the signal. The decomposition accuracy is mainly controlled by the

penalty factor α and the modal decomposition number k. A large k value will cause the signal to be over-decomposed, while a too small k may lead to incomplete decomposition [33], resulting in insufficient extraction of noise. The penalty factor α determines the bandwidth of each modal component, and different bandwidth scales have an impact on the signal extraction results. Since the landslide displacement time series is complex and changeable, it is difficult to select appropriate k and α, and it is easy to increase the randomness of the decomposition results.

In order to solve these problems, this paper uses CLF-SSA to optimize the parameter selection of VMD to achieve a more thorough signal decomposition. The size of the envelope entropy Ep is closely related to the sparsity of the signal. The smaller the envelope entropy value, the more ordered the Intrinsic Mode Functions (IMF), the higher the sparsity, and the less noise [34], which is more conducive to subsequent prediction. Therefore, this paper selects the minimum envelope entropy as the fitness parameter.

The envelope entropy calculation formula of sequence $x(i)$ is:

$$\begin{cases} E_p = -\sum_{i=1}^{N} p_i \lg p_i \\ p_i = \dfrac{a(j)}{\sum_{i=1}^{N} a_i} \end{cases} \tag{12}$$

where $j = 1, 2, \cdots, N$, N is the number of sampling points, $a(j)$ is the envelope signal obtained after Hilbert transform of $x(j)$, and $p(j)$ is the normalized form of $a(j)$.

The minimum envelope entropy calculation formula of the fitness function is:

$$fitness(VMD(k, \alpha)) = \frac{1}{K} \sum_{i=1}^{k} E_p(i) \tag{13}$$

SSA is used to optimize the penalty factor α and modal decomposition number k of VMD, and the envelope entropy is used as the fitness value. The specific steps are as follows:

Step 1: Set the range of parameters k and α;

Step 2: Initialize the parameters, use VMD to decompose the landslide displacement, and set the envelope entropy as fitness;

Step 3: Update the individual position according to the con-ditions of the individual's current environment;

Step 4: Determine whether the current fitness is greater than the previous fitness. If so, update the fitness. If not, keep it unchanged;

Step 5: Determine whether the algorithm termination condi-tion is met. If so, output the optimal parameters and fitness. Otherwise, return to step 2.

C. ARIMA

ARIMA model is an improved model of the Autoregressive Moving Average (ARMA) model. The ARMA model ass-umes that the data is stationary when making predictions, that is, the mean, variance and autocorrelation function of the data do not change with time. The model has a better prediction effect in stationary series. If the data is non-stationary, false regression will occur [35].

In order to avoid the occurrence of false regression, ARIMA processes non-stationary series through the difference method, converts it into a stationary series, and uses ARMA for prediction. There are three parameters p, d, and q in ARIMA. p represents the autoregressive order, d represents the difference order required for the non-stationary series, and q represents the moving average order. The ARIMA *(p,d,q)* formula is as follows:

$$X_t = \sum_{i=1}^{p} \varphi_i X_{t-1} + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} + \varepsilon_t \tag{14}$$

where φ_i represents the i-th order autoregressive coefficient, θ_j represents the j-th order moving average coefficient, ε_t represents the t-th order error term.

The ARIMA model uses historical data to predict future values. It has high prediction accuracy, a simple model, and only requires endogenous variables without the need for other exogenous variables [36]. Therefore, this paper uses the ARIMA model to predict landslide trend items, which can effectively improve the efficiency of landslide prediction.

D. BiLSTM-ATTENTION COMBINATION MODEL

1) LSTM

LSTM is a deep learning model for processing complex nonlinear time series. Through the mechanism of gated units, it deeply explores the spatial relationship between time series and each moment. LSTM is an improvement of RNN. In long time series analysis, RNN has problems such as gradient vanishing and gradient explosion during the back propagation process [37], which makes the model unable to obtain information about the forward sequence and makes the training unstable. In order to improve such problems, LSTM proposes the concepts of forget gate, input gate, output gate and memory unit [38]. The forget gate retains the information of the previous moment, compares it with the current input information, and uses the Sigmoid function to map and retain part of the information. The input gate uses the Sigmoid function and the Tanh function to generate the information that needs to be added to the memory unit. The output gate passes the processed current information to the next moment. The memory unit is the core of LSTM. Through the combination of forget gate and input gate data processing, the important information of each moment is retained and the unimportant information of each moment is forgotten. The calculation formula is as follows:

$$f_t = \sigma\left(w_f \cdot [h_{t-1}, x_t] + b_f\right) \tag{15}$$

$$i_t = \sigma(w_i \cdot [h_{t-1}, x_t] + b_i) \tag{16}$$

$$o_t = \sigma(w_o \cdot [h_{t-1}, x_t] + b_o) \tag{17}$$

$$C_t = f_t \cdot C_{t-1} + i_t \cdot C'' \tag{18}$$

$$h_t = o_t \cdot \tan\ h(C_t) \tag{19}$$

$$\tilde{C}_t = \tan h(W_c \cdot [h_{t-1}, x_t] + b_c) \tag{20}$$

where x_t is the input data at the current moment, f_t is the forget gate, i_t is the input gate, O_t is the output gate, h_{t-1} is the output data at the previous moment, h_t is the output data at the current moment, C_t is the memory unit, w_f,w_i,w_o are the weights of the forget gate, input gate and output gate respectively, b_f,b_i,b_o are the biases of the forget gate, input gate and output gate respectively, and the model framework is shown in Fig 2.

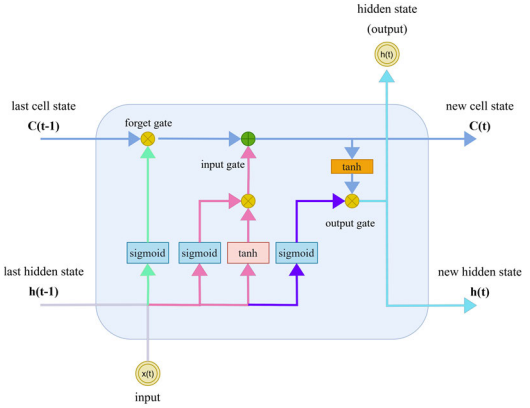


FIGURE 2. LSTM model framework.

LSTM uses gating mechanisms and memory unit information to deeply mine the temporal and spatial nonlinear relationships between data, capture the complex dependencies between data, and has higher stability and prediction capabilities. At the same time, due to the multiplication method of each hidden unit in the traditional RNN, there are problems of gradient explosion and gradient vanishing. LSTM introduces a nonlinear change method to transmit information [39], which to a certain extent solves the problems of gradient explosion and gradient vanishing in long time series.

2) BILSTM

The LSTM model explores the spatiotemporal relationship between data and time through forward information transfer, thus improving the model accuracy. In BiLSTM, the backward information transfer LSTM method is introduced to fully consider the dependence of current data on previous and next information, so as to better adapt to different data in complex time series and improve the robustness and generalization ability of the model [40].

BiLSTM introduces two layers of forward and backward LSTM networks to recombine the current output information. Through the information splicing method, the information obtained at the current moment is made more comprehensive, the information adaptability is improved, and the model can better understand the complex nonlinear relationship in the time series [41]. The calculation formula is as follows:

$$\overrightarrow{H_f} = \overrightarrow{LSTM}(h_{t-1}, x_t, c_{t-1}), \quad t \in [1, T] \tag{21}$$

$$\overleftarrow{H_b} = \overleftarrow{LSTM}(h_{t+1}, x_t, c_{t+1}), \quad t \in [T, 1] \tag{22}$$

$$H_t = \left[\overrightarrow{H_f}, \overleftarrow{H_b}\right] \tag{23}$$

where $\overrightarrow{H_f}$ is the information obtained by the forward LSTM, $\overleftarrow{H_b}$ is the information obtained by the backward LSTM, H_t is the information at time t, and the model structure is shown in Fig 3.

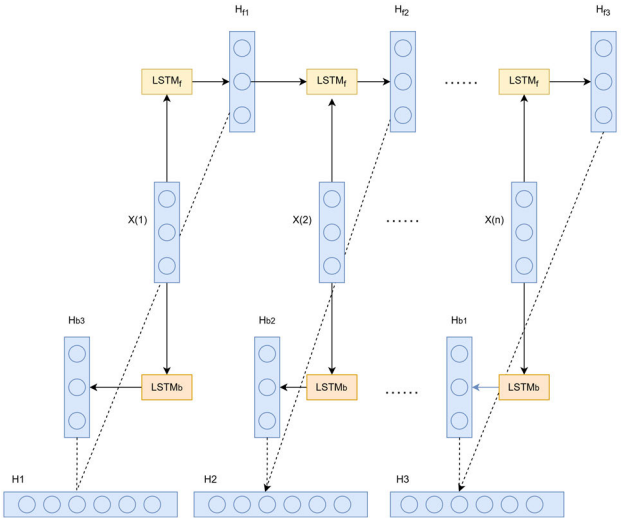


FIGURE 3. BiLSTM model framework.

3) ATTENTION MECHANISM

When the model processes a large amount of information, the dimension of the input information determines the time required for model training. In the field of landslide displacement prediction, prediction efficiency is very important. In order to improve the efficiency of the prediction model, the Attention mechanism is introduced to extract key features. By calculating the contribution value of the input feature to each position, a weight parameter is assigned to each input feature [42]. The calculation formula is as follows:

$$c^t = \sum_{i=1}^{T} w^{ti} \cdot h^i \tag{24}$$

$$y^t = m(c^t, g^{t-1}, y^{t-1}) \tag{25}$$

where c^t is all weighted semantic features, w is the attention weight, h is the encoding result, y is the output information, g^{t-1} is the hidden state at the previous moment, and the model structure is shown in Fig 4.

Traditional time series prediction methods suffer from the phenomenon of forward information decay. The Attention mechanism can focus on the historical information at each moment and easily obtain the data dependency in long time series. At the same time, the Attention mechanism can dynamically adjust parameters and has a good effect in processing non-stationary series. Since the Attention mechanism has no information dependency in the time step [43], it can

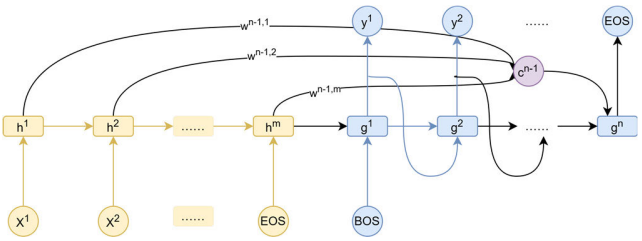


FIGURE 4. Attention Mechanism framework.

calculate information in parallel, which greatly improves the prediction efficiency and accuracy of the model in long time series.

E. GREY RELATIONAL ANALYSIS

Grey Relational Analysis (GRA) is used to explore the degree of correlation between two sequences. In mathematical statistics, regression analysis, variance analysis, principal component analysis and other methods are often used to study similar problems, but these methods usually require a large amount of data and require the data to meet certain distribution characteristics. In landslide displacement prediction, in order to improve prediction efficiency, a large amount of data is often not used for training [44]. Grey correlation analysis does not require the distribution characteristics or magnitude consistency of sample data, but judges the correlation by comparing the morphology of the data sequence, so it can effectively solve such problems. The steps are as follows:

Step 1, assuming that the original data sequences x_0 and x_1 are:

$$\begin{cases} x_0 = (x_0(1), x_0(2), \cdots, x_0(n) \\ x_1 = (x_1(1), x_1(2), \cdots, x_1(n) \end{cases} \tag{26}$$

Step 2: preprocess the data and standardize each element by dividing the mean of the sequence, that is:

$$x_{ij} = \frac{x_{ij}}{\frac{1}{n}\sum_{i=1}^{n} x_{ij}} \tag{27}$$

Step 3, calculate the minimum difference a and maximum difference b of sequences x_0 and x_1.

$$\begin{cases} a = \min |x_0(k) - x_1(k)| \\ b = \max |x_0(k) - x_1(k)| \end{cases} \tag{28}$$

Step 4: Finally, calculate the correlation value of the two sequences.

$$\gamma(x_0(k), x_1(k)) = \frac{a + pb}{|x_0(k) - x_1(k) + pb|} \tag{29}$$

where p is the resolution coefficient, the value range is [0,1], and it is usually 0.5. When $p = 0.5$, $\gamma(x_0(k), x_1(k)) > 0.6$, it is considered that the two sequences are correlated, and the $\gamma(x_0(k), x_1(k))$ larger the stronger the correlation.

F. PERFORMANCE EVALUATION METRICS

In order to evaluate the results of the prediction model, the Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and R^2 were used to judge the true value and the predicted value [45].

RMSE is the sum of the squares of the deviations between the predicted value and the true value, and the calculation formula is:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2} \tag{30}$$

MAE is the average deviation between the predicted value and the true value, and the calculation formula is:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|\hat{y}_i - y_i| \tag{31}$$

R^2 represents the correlation between the predicted value and the true value, ranging from [0,1]. The larger the value, the closer the correlation. The calculation formula is:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{\sum_{i=1}^{n}(\hat{y}_i - y_i)} \tag{32}$$

where: $\hat{y}_i$ is the true value, y_i is the predicted value, and n is the number of data.

G. LANDSLIDE DISPLACEMENT PREDICTION PROCESS

This paper designs a landslide displacement prediction model based on optimization decomposition and deep attention mechanism, proposes Chebyshev chaos map initialization population strategy and Levy flight strategy to improve the optimization ability of SSA, introduces CLF-SSA-VMD algorithm to deeply decompose and reconstruct landslide displacement, extracts landslide trend terms and periodic terms, and uses ARIMA model to fit trend terms to improve prediction efficiency. In order to improve prediction accuracy, BiLSTM-Attention dynamic prediction model is designed to deeply explore the dependency between landslide displace-ment sequence and influencing factors. The process design is shown in Fig 5, and the implementation steps are as follows:

Step 1: For the landslide displacement sequence, CLF-SSA is introduced to optimize the VMD algorithm parameters and find the optimal decomposition parameters, and then each landslide displacement sequence is reorganized into trend items and period items;

Step 2: The ARIMA model and grid search algorithm are introduced to predict the trend item to verify the prediction ability and decomposition accuracy;

Step 3: Eight alternative influencing factor sequences are proposed, and the grey correlation method is used to compare them with the landslide period item to screen the influencing factor sequence with higher correlation;

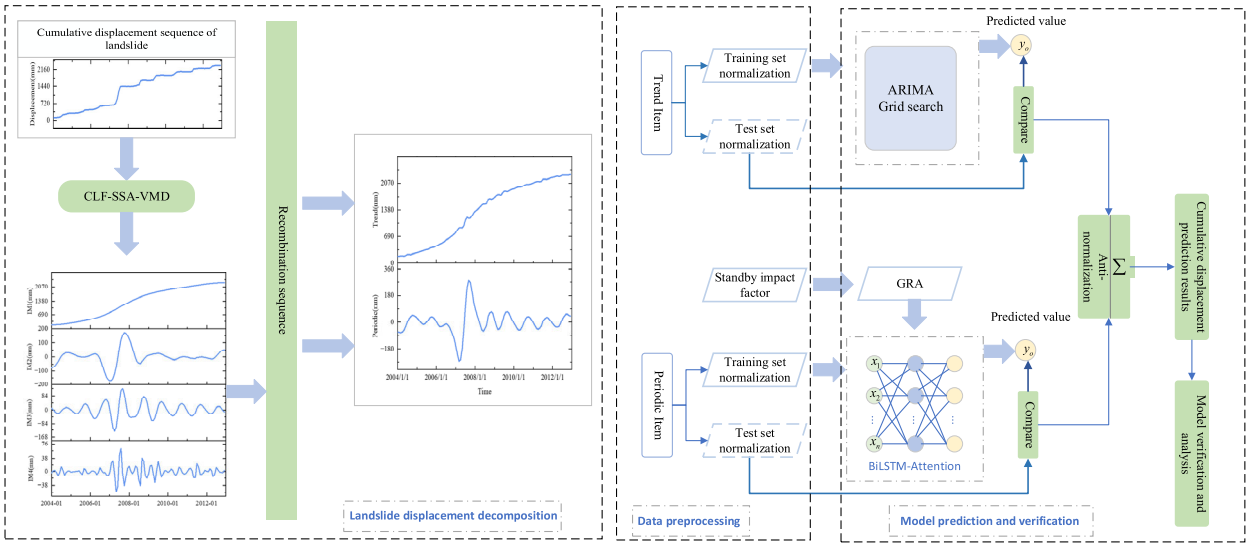


FIGURE 5. Landslide displacement prediction process.

Step 4: The BiLSTM-Attention dynamic prediction model is designed and applied to the prediction of landslide displacement period item. At the same time, four horizontal comparison models are used to conduct experimental tests under the same input conditions to verify the prediction effect of the BiLSTM-Attention model.

Step 5: Finally, four landslide displacement prediction models are introduced for experimental comparison with the prediction model proposed in this paper, and the effectiveness of the model in this paper is verified by three indicators: RMSE, R^2 and MAE.

III. BAISHUI RIVER LANDSLIDE SITUATION

The Baishuihe landslide is located on the southern bank of the Yangtze River, approximately 56 kilometers from the Three Gorges Dam, in Shazhenxi Town, Zigui County, Hubei Province. The landslide body comprises an accumulation layer structure, characterized by a stepped distribution descending toward the Yangtze River, with lower elevations in the north and higher elevations in the south. The landslide measures approximately 600 meters from south to north, 700 meters from east to west, with a volume of about 12.6 million cubic meters and an average thickness of 30 meters. The rear edge of the landslide has an elevation of about 410 meters [46], while the front edge lies beneath the Yangtze River reservoir water level, ranging from approximately 70 to 145 meters in elevation. The main sliding direction is approximately 20°, with movement along a monocline slope. The rear edge is steep, while the front edge is relatively gentle.

A total of 11 GPS monitoring points have been installed, as illustrated in Fig 6. Seven of these points are distributed along three longitudinal profiles, with three located on the central profile and two on each of the side profiles. The remaining four monitoring points are positioned on the bedrock ridges to the east and west of the landslide warning area. Additionally, a GPS reference point is established on each of the four profiles. The monitoring system primarily measures surface displacement, borehole inclinometer data, and groundwater levels [47].

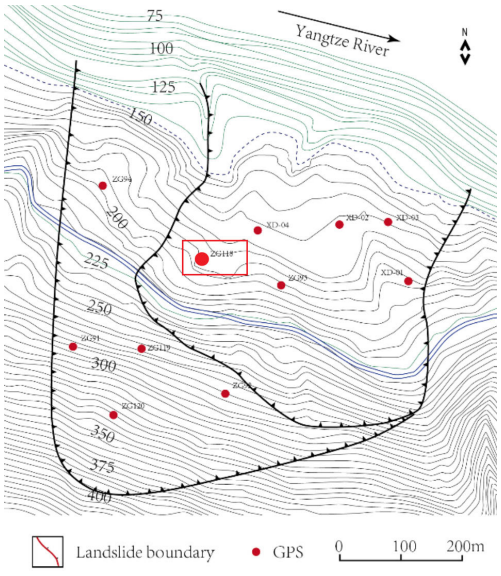


FIGURE 6. GPS installation positions.

Among these, the Z118 monitoring point, situated in the central part of the landslide, has the longest monitoring history and effectively captures the landslide's evolution. This study utilizes cumulative displacement data from the Z118 monitoring point, covering the period from January 2004 to December 2012. The data, recorded monthly, includes cumulative rainfall, average reservoir water levels, and monthly landslide displacement, as shown in Fig 7. These datasets

were sourced from the National Glacier, Frozen Soil, and Desert Data Center [48].

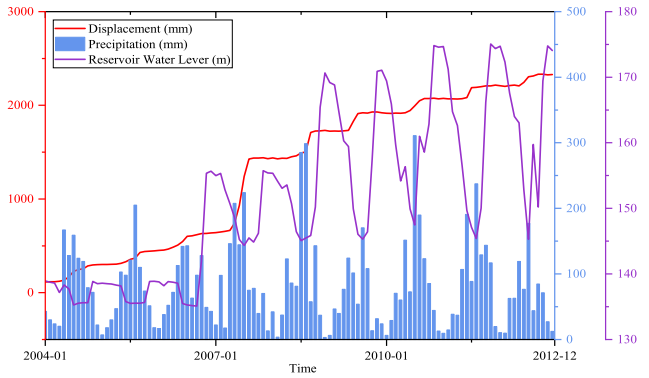


FIGURE 7. Displacement and environmental data variation in the Baishuihe landslide.

During the rainy season (May to September), significant rainfall increases soil moisture in the landslide area, softens the soil, reduces its stability, and raises the risk of landslides. As illustrated in the figure, landslide displacement correlates strongly with rainfall, displaying a seasonal variation pattern. Conversely, during the dry season (October to April), reduced rainfall stabilizes the landslide displacement [49].

The reservoir water level fluctuates cyclically between 145 and 175 meters, exerting a substantial influence on landslide stability. When the reservoir water level decreases, uneven water pressure inside and outside the landslide body alters the pore water pressure gradient, generating an unstable seepage field and increasing landslide displacement. Conversely, rising water levels lead to infiltration into the landslide body, reducing soil cohesion and internal friction angles, thereby weakening stability and exacerbating sliding tendencies [50].

Following heavy rainfall, landslide displacement often continues to increase due to the retention of surface water, leaving the landslide body in a moist, unstable state. Furthermore, during periods when the reservoir water level reaches its lowest point, the landslide displacement growth rate accelerates significantly. Both rainfall and reservoir water level changes exhibit a lagged impact on landslide displacement, highlighting the necessity of accounting for this hysteresis effect when analyzing influencing factors.

IV. CLF-SSA ALGORITHM PERFORMANCE TEST

In order to verify the performance of the improved CLF-SSA optimization algorithm in this paper, this paper introduces 8 international standard test functions to verify the performance [51], as shown in Fig 8 and Table 1. F1∼F4 are unimodal test functions, and F5∼F8 are multimodal test functions. The unimodal test function verifies the algorithm's convergence speed and local search capability, and the

TABLE 1. Test functions.

Type	Name	Function	Dim	Range	F_{min}
Unimodal test functions	Sphere	$F_1(x)=\sum_{i=1}^{n}x_i^2$	30	[-100,100]	0
	Schwefel 2.22	$F_2(x)=\sum_{i=1}^{n}\lvert x_i\rvert+\prod_{i=1}^{n}\lvert x_i\rvert$	30	[-10,10]	0
	Schwefel 1.2	$F_3(x)=\sum_{i=1}^{n}(\sum_{j-1}^{i}x_j)^2$	30	[-100,100]	0
	Schwefel 2.21	$F_4(x)=\max_i\{\lvert x_i\rvert,1\le i\le n\}$	30	[-100,100]	0
Multimodal test functions	Rastrigin	$F_5(x)=\sum_{i=1}^{n}[x_i^2-10\cos(2\pi x_i)+10]$	30	[-5.12,5.12]	0
	Ackley	$F_6(x)=-20\exp\left(-0.2\sqrt{\frac{1}{n}\sum_i^n x_i^2}\right)-\exp\left(\sqrt{\frac{1}{n}\sum_i^n 2\pi x_i}\right)+20+e$	30	[-32,32]	0
	Griewank	$F_7(x)=\frac{1}{4000}\sum_{i=1}^{n}x_i^4+\prod_i^n\cos\left(\frac{x_i}{\sqrt{i}}\right)+1$	30	[-600,600]	0
	Shekel's Family	$F_8(x)=-\sum_{i=1}^{10}[(x-a_i)(x-a_i)^T+c_i]^{-1}$	4	[0,10]	-10.5363

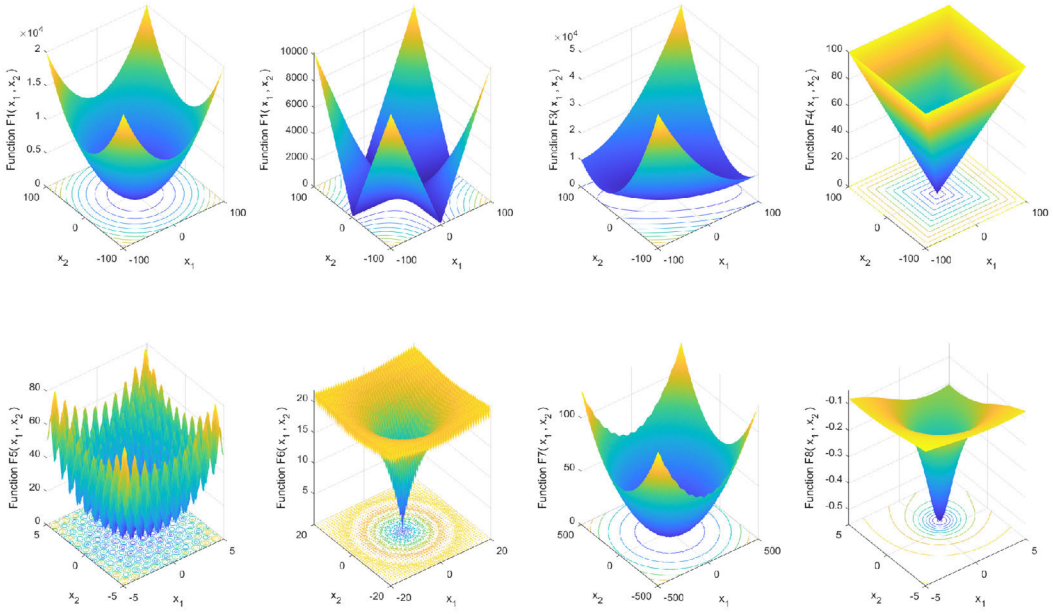


FIGURE 8. Test functions.

TABLE 2. Comparison of benchmark function test results of four optimization algorithms.

Function	GWO		GJO		SSA		CLF-SSA	
	Avg	Std	Avg	Std	Avg	Std	Avg	Std
F1	6.75E-28	1.98E-27	1.14E-54	3.95E-53	1.66E-55	5.71E-63	**0**	**0**
F2	8.72E-17	7.86E-17	2.87E-32	3.97E-32	6.37E-30	8.61E-32	1.69E-191	**0**
F3	5.56E-06	4.14E-05	9.42E-17	1.65E-16	1.12E-26	1.77E-27	**0**	**0**
F4	5.8E-07	7.24E-07	1.12E-15	2.56E-14	3.12E-33	6.12E-29	1.85E-191	**0**
F5	2.639176	3.777656	**0**	**0**	**0**	**0**	**0**	**0**
F6	1.01E-13	1.64E-14	6.36E-15	1.59E-15	4.44E-16	**0**	4.44E-16	**0**
F7	0.004507	0.007450	0.000293	**0**	**0**	**0**	**0**	**0**
F8	-10.35448	0.001211	-10.07731	1.362392	-9.582863	2.592884	-10.53641	**4.08E-15**

multimodal test function verifies the algorithm's global search capability.

Three popular intelligent optimization algorithms, GWO [51], GJO [52], and SSA [28], are introduced for experimental comparison with CLF-SSA. Since the optimization algorithms are all random, to ensure the validity of the data, this paper runs the four algorithms 30 times and iterates 600 times respectively, and takes the mean and standard deviation of the results. The mean represents the optimization ability of the algorithm, and the standard deviation represents the stability of the algorithm. A set of optimization processes are randomly selected as shown in Fig 9, and the mean and standard deviation results are shown in Table 2.

In the unimodal test function, CLF-SSA showed strong optimization ability and efficiency. Compared with the original SSA algorithm, it can quickly find the optimal solution in the early stage of iteration, effectively improving the convergence speed. In the multimodal test function, CLF-SSA chaotic mapping initializes the population, the population is diversified, and the efficiency and accuracy are the fastest when seeking the optimal solution. In the F8 test function, all four algorithms fell into the local optimal solution in the early stage of iteration. The CLF-SSA algorithm jumped out of the local optimal solution the fastest through the Levy random step, and calculated the global optimal solution in less than 50 iterations. This result shows that the chaotic mapping initialized population and the Levy flight improved SSA algorithm have achieved good results.

V. LANDSLIDE DISPLACEMENT PREDICTION

This study employs data from the ZG118 monitoring point of the Baishui River landslide as the research object. The dataset spans January 2004 to December 2012, with monthly records comprising a total of 108 data points. The cumulative landslide displacement at this monitoring point is 2215 mm. The data is divided into a training set (January 2004 to February 2011) for model training and a test set (March 2011 to December 2012) for model validation.

A. CLF-SSA-VMD DECOMPOSITION DISPLACEMENT

In order to decompose the displacement more thoroughly and improve the accuracy and generalization ability of the landslide displacement prediction model, CLF-SSA is introduced to find the optimal parameters k and α of the VMD

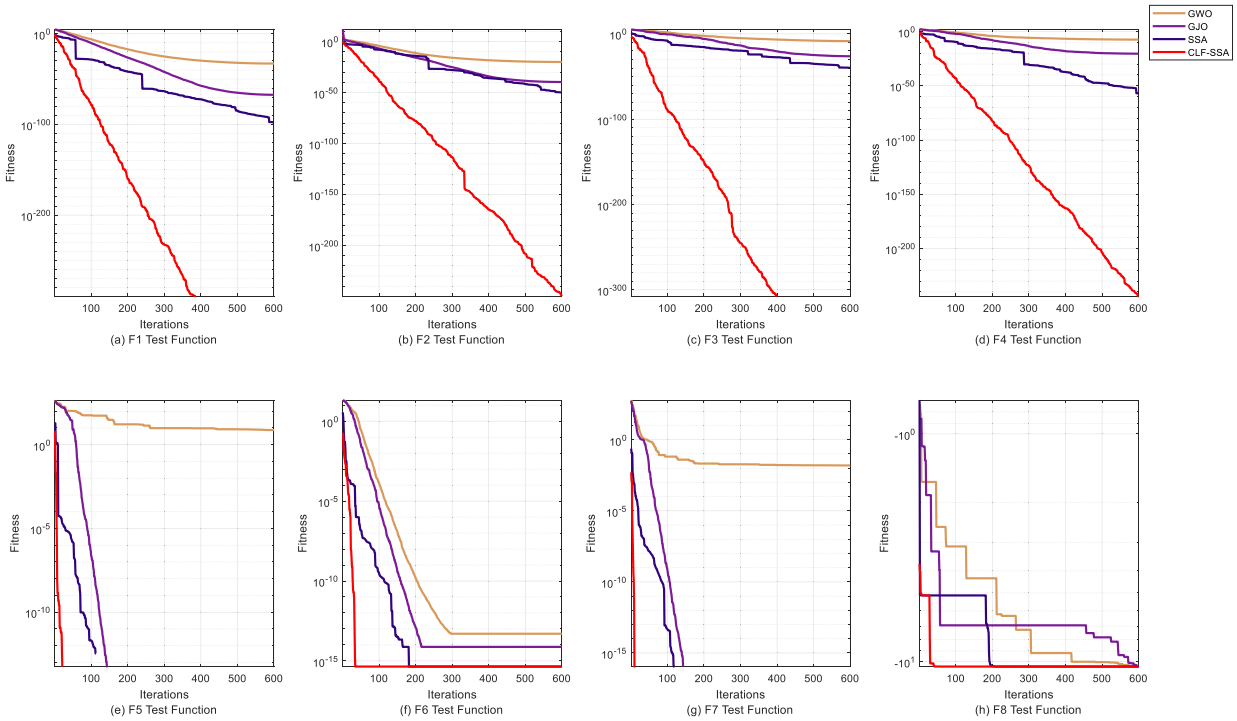


FIGURE 9. Optimization and iterative effects of four types of optimization algorithms.

decomposition algorithm. In the CLF-SSA algorithm, the population size is set to 30, the number of iterations is 20, the decomposition parameter k is set in the range of [2], [10], and the penalty factor parameter α is set in the range of [1,3000].

After multiple iterations, the optimal result is $k = 4$, $\alpha = 26$. The minimum envelope entropy value of the fitness function is 4.1725. The optimal parameters are brought into the VMD algorithm to decompose the landslide displacement. The decomposition result is shown in Fig 10.

After decomposition, four IMF components are obtained, among which IM1 has obvious trend, IM2 and IM3 have obvious periodicity, and IM4 has large fluctuations because there are random fluctuations in the landslide process. Since the generation of random fluctuations includes uncertain factors such as human activities and is difficult to predict, this paper reorganizes IM1 and IM4 into trend items, and IM2 and IM3 into periodic items. The reorganization results are shown in Fig 11.

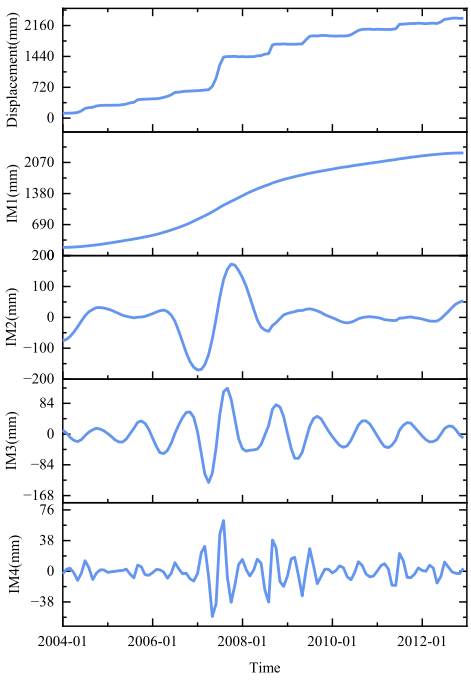


FIGURE 10. Landslide displacement decomposition.

B. TREND ITEM DISPLACEMENT PREDICTION

The trend term represents the long-term evolution of the landslide. After being decomposed by the VMD algorithm, the trend term is smoother. In order to improve the efficiency of the landslide displacement prediction model, this paper introduces the ARIMA model for trend term prediction and uses the grid search method to optimize the parameters of ARIMA. The optimal parameters are $p = 3$, $d = 1$, $q = 2$. The prediction results are shown in Fig 12, with R^2 of 0.99, RMSE of 6.11mm, MAE of 5.41mm, maximum error of 11.43mm, and minimum error of 0.22mm. The fitting effect is good.

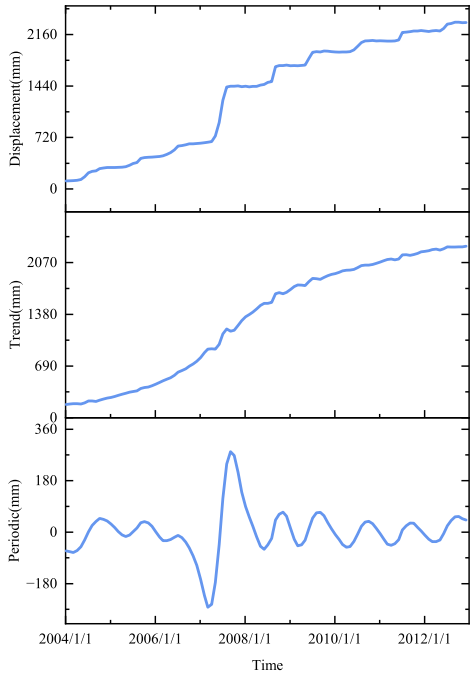


FIGURE 11. Landslide displacement reconstruction.

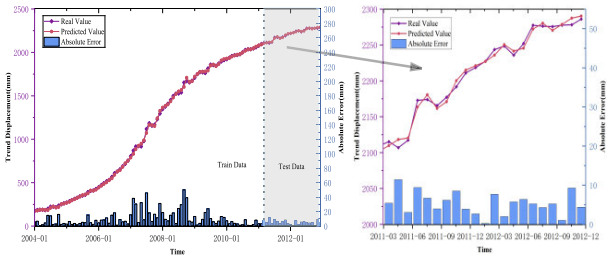


FIGURE 12. Fitting effect of landslide trend term.

C. PERIODIC TERM DISPLACEMENT PREDICTION

The periodic term represents the process of landslide evolution with seasonal cycle changes. According to the analysis in Chapter III, this paper proposes three types of factors as alternative influencing factors, namely: rainfall, reservoir water level and landslide displacement. The rainfall category proposes the cumulative rainfall of the current month P_1, the cumulative rainfall of the previous month P_2 and the rainfall change between two months P_3 as alternative influencing factors. The reservoir water level category proposes the average reservoir water level of the current month R_1, the average reservoir water level of the previous month R_2 and the average reservoir water level change between two months R_3 as alternative influencing factors. The landslide displacement category proposes the landslide displacement of the previous month D_1 and the landslide displacement change of the previous month D_2 as alternative influencing factors.

Selecting influencing factors with high correlation can improve the accuracy of landslide displacement prediction. In order to screen accurate input influencing factor sequences, GRA is introduced to judge the correlation between the influencing factor sequence and the periodic term. The grey correlation resolution parameter is taken as 0.5. If the calculated correlation result is greater than 0.6, the two sequences can be considered to be correlated. The results are shown in Fig 13.

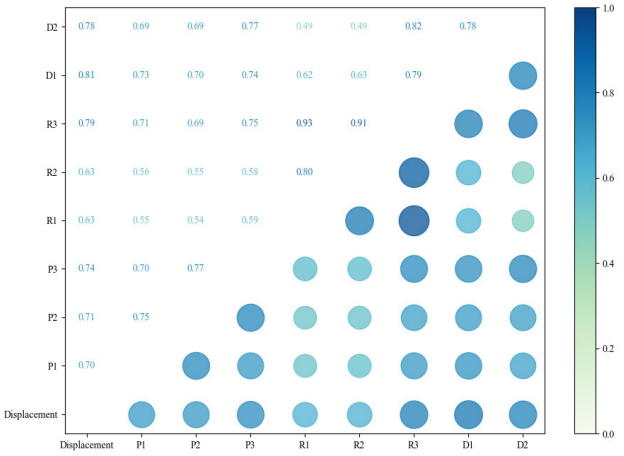


FIGURE 13. Grey correlation value of impact factor.

As can be seen from the figure above, each candidate influencing factor is greater than 0.6 and can be used as an input variable. This is because the CLF-SSA-VMD decomposition algorithm effectively removes the noise sequence in the landslide displacement period term. However, when there are many input influencing factors, the prediction model needs to be mapped to a high-dimensional space, resulting in low prediction efficiency. For this reason, this paper selects P_3, R_3, D_1 and D_2 with high correlation as input variables to improve the speed and accuracy of the prediction model.

The BiLSTM-Attention prediction model is used for periodic item prediction. A two-layer BiLSTM neural network is constructed. The number of neurons is 128, the Dropout layer parameter is 10%, the Adam method is selected as the optimization function, and the MAE is selected as the loss function. The results of 100 iterations show that the R^2 is 0.99, the RMSE is 2.09mm, and the MAE is 1.83mm. The prediction results are shown in Fig 14.

In order to verify the accuracy of the BiLSTM-Attention prediction model, BiGRU, BiLSTM, LSTM-Attention, and GRU-Attention models were introduced for experimental comparison. The model parameters are the same as those of BiLSTM-Attention. The experimental results are shown in Fig 15 and Table 3.

The effect of BiLSTM-Attention is the best in both the training set and the test set, because BiLSTM can extract current information from the future and construct reverse features to improve the model prediction accuracy. The Attention mechanism effectively backpropagates the model input variables, constructs the weights of the input variables, and improves the prediction accuracy by weighting.

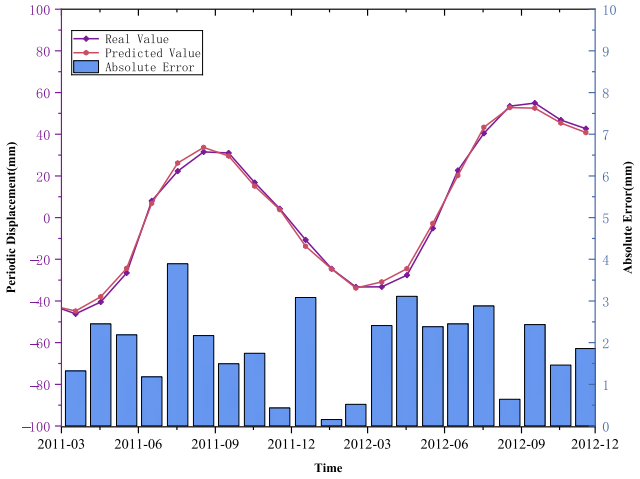


FIGURE 14. Fitting effect of landslide period term.

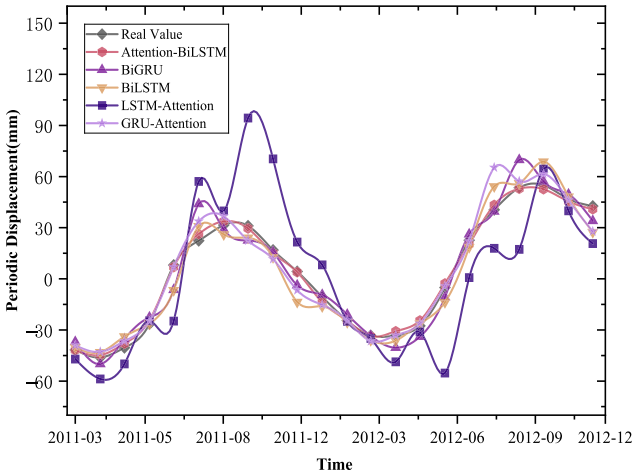


FIGURE 15. Comparison of the effects of 5 prediction models.

TABLE 3. Comparison of test and validation sets of 5 prediction models.

Model	Type	R^2	RMSE(mm)	MAE(mm)
BiLSTM-Attention	Train	**0.98**	**12.93**	**8.83**
	Test	**0.99**	**2.09**	**1.83**
BiGRU	Train	0.91	25.76	17.26
	Test	0.94	7.99	6.03
BiLSTM	Train	0.94	21.39	14.09
	Test	0.93	8.33	6.47
LSTM-Attention	Train	0.55	58.93	36.97
	Test	0.35	26.86	20.41
GRU-Attention	Train	0.96	16.01	11.20
	Test	0.94	7.82	5.25

D. CUMULATIVE DISPLACEMENT PREDICTION

According to the time series addition model, the cumulative displacement prediction result is equal to the sum of the trend term and the period term prediction results. The prediction results are shown in Fig 16. The R^2 value is 0.99, the RMSE value is 5.91mm, and the MAE value is 5.10mm, which accurately reflects the future trend of the landslide.

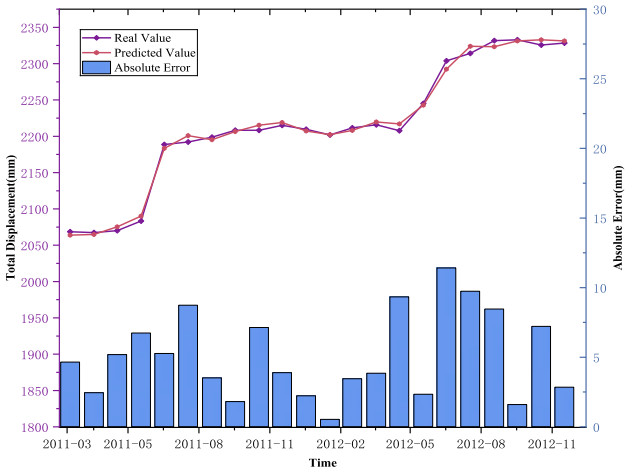


FIGURE 16. Fitting effect of total landslide displacement.

To further assess the predictive performance of the proposed model, we introduce the Mean Absolute Scaled Error (MASE) as an additional evaluation metric. Unlike conventional error measures such as RMSE and MAE, MASE provides a scale-independent assessment by comparing the model's performance against a simple baseline forecast. A MASE value below 1 indicates that the proposed model outperforms the baseline, whereas a value above 1 suggests otherwise.

In this study, we compare the proposed model with two baseline approaches: the Naïve Persistence Model, which assumes that the displacement remains unchanged in the next time step, and the Moving Average Model, which predicts displacement based on the mean of the previous three observations. The experimental results show that the Naïve Persistence Model achieves an MAE of 15.21 mm, while the Moving Average Model, with a window size of three, yields an MAE of 25.18 mm, indicating that simple averaging fails to capture the nonlinear characteristics of landslide displacement. In contrast, the proposed model significantly reduces the prediction error, achieving an MAE of 1.83 mm. Consequently, the MASE of the proposed model, calculated using the Naïve Persistence Model as the baseline, is 0.12, confirming its superior predictive capability. These results highlight the limitations of traditional statistical approaches in handling the complex temporal dependencies of landslide displacement, whereas the proposed model, leveraging adaptive decomposition and deep learning, provides more accurate and reliable forecasts.

E. MODEL PREDICTION EFFECT VERIFICATION

In order to further verify the effect of the model, this paper introduces XgBoost, random forest RF, EMD-BiLSTM, and EEMD-GRU landslide displacement prediction models for experimental comparison. The EMD and EEMD

decom-position parameters use the default values, the number of GRU and BiLSTM neurons is 128, the number of layers is 2, and the Dropout layer is 10%. The RMSE, R^2, and MAE indicators are used to evaluate the prediction model. The results are shown in Fig 17 and Table 4.

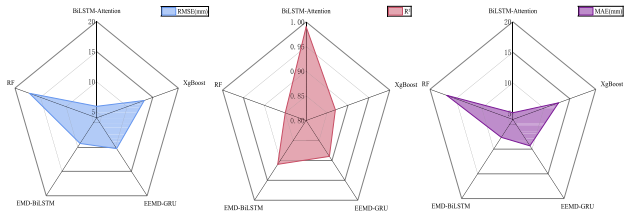


FIGURE 17. Comparison of evaluation indicators of 5 prediction models.

TABLE 4. Comparison of evaluation indicators of 5 prediction models.

Model name	RMSE	MAE	R^2
XgBoost	12.81	13.32	0.87
RF	16.78	17.18	0.85
EMD-BiLSTM	7.62	9.27	0.91
EEMD-GRU	9.33	10.33	0.89
Proposed Model	**5.1**	**5.91**	**0.99**

The data show that the prediction model proposed in this paper has a good effect. Compared with the traditional machine learning XgBoost and RF prediction models, the RMSE is reduced by 60% and 69% respectively, and the R^2 is increased by 13% and 16% respectively. Compared with the traditional decomposition algorithm combined with the neural network EMD-BiLSTM and EEMD-GRU prediction models, the RMSE is reduced by 33% and 45% respectively, and the R2 is increased by 8% and 11%. Because the CLF-SSA-VMD decomposition algorithm decomposes the displacement more thoroughly, it improves the generalization ability and robustness of the model. The BiLSTM-Attention prediction model introduces bidirectional information transmission and weight mechanism to improve the prediction ability of the model.

To further evaluate the computational efficiency of the proposed model, we compare its execution time with existing models. To ensure the reliability of the evaluation, each model was executed 20 times, and the average runtime was recorded. All experiments were conducted on a high-performance workstation equipped with an Intel Core i9-13900K CPU, an NVIDIA GeForce RTX 4090 GPU, and a software environment including Python 3.9, PyTorch 2.0, and TensorFlow 2.10 with CUDA acceleration.

Experimental results indicate that the proposed model requires an average computation time of 5.98 seconds per prediction. In contrast, EMD-BiLSTM and EEMD-GRU take 4.93 and 5.18 seconds, respectively, while traditional machine learning models such as XGBoost and Random Forest (RF) complete the prediction in 4.20 and 3.73 seconds, respectively.

The increased computational cost of the proposed model is primarily attributed to the incorporation of VMD decompo-sition and the BiLSTM-Attention architecture, which enhance feature extraction and temporal dependency modeling. However, the CLF-SSA optimization significantly improves parameter selection efficiency, reducing redundant comput-ations and enhancing model generalization. Despite its slightly higher computational overhead, the proposed model demonstrates superior predictive performance, justifying the trade-off between accuracy and efficiency in landslide displacement forecasting.

Each predictive model has certain advantages and limitations, making them more suitable for specific forecasting scenarios. Traditional machine learning models, such as XGBoost and RF, offer faster computation and are well-suited for applications where real-time predictions and low compu-tational costs are priorities. However, they often struggle to capture complex temporal dependencies and nonlinear patterns, leading to reduced accuracy in long-term landslide displacement forecasting.

On the other hand, deep learning-based models, particularly those incorporating sequence modeling techniques like BiLSTM, are more adept at capturing long-term dependencies and intricate temporal patterns. These models perform well in landslide displacement forecasting tasks where displacement trends exhibit strong seasonality, multi-scale variations, and long-range dependencies. However, their increased comple-xity and training costs make them less suitable for real-time applications with strict latency constraints or resource-limited environments.

Furthermore, the integration of VMD enhances the model's ability to handle multi-scale displacement variations and mitigate noise effects, thereby improving predictive accuracy. However, this also increases the computational burden, making the proposed model more suitable for high-precision forecasting tasks, such as early warning systems and long-term landslide monitoring, rather than instant real-time applications.

Despite its advantages, the proposed model has certain limitations. First, the computational overhead remains higher than traditional machine learning approaches, which may restrict its deployment in environments with limited computational resources. Second, the model's performance relies on the effectiveness of the VMD decomposition, meaning that its accuracy could be affected by inappropriate parameter settings. Although CLF-SSA helps optimize the decomposition process, further research is needed to explore adaptive decomposition techniques to further enhance efficiency and robustness.

In summary, while the proposed model achieves higher predictive accuracy by leveraging advanced decomposition techniques and deep learning architectures, it comes at the cost of increased computational complexity. Thus, it is best suited for landslide displacement forecasting tasks that require high accuracy and can accommodate moderate computational costs, such as disaster prevention and risk assessment applications. For real-time applications with strict efficiency constraints, traditional machine learning

approaches remain a viable alternative, albeit with a potential compromise in prediction accuracy.

VI. CONCLUSION

In light of the constantly shifting and nonlinear nature of landslide displacement prediction, this investigation suggests a predictive framework that combines deep learning and optimum decomposition. Three solutions are shown for enhancing the generalization and application of SSA in order to boost its overall effectiveness. In order to reduce the need for human parameter selection and more efficiently extract landslide displacement components, SSA is further incorporated with VMD.

A BiLSTM-Attention model is created to suit the periodic term, and ARIMA is used to represent the trend term in order to simulate various displacement components. Eight other influencing elements are also added as input variables, which greatly raises the model's predicted accuracy.

Future research should incorporate more landslide-inducing factors, such as wind speed, temperature, humidity, vegetation coverage, and human activities, to enhance the interpretability of the trend term and develop a multi-source landslide displacement prediction framework. Furthermore, since the influence of rainfall on landslide displacement exhibits a lag effect that varies across regions, future studies should explore region-specific rainfall hysteresis periods to improve the interpretability of periodic displacement variations.

REFERENCES

[1] G. Shanmugam and Y. Wang, "The landslide problem," *J. Palaeogeography*, vol. 4, no. 2, pp. 109–166, Apr. 2015.

[2] *Ministry of Natural Resources of the People's Republic of China*, China Natural Resour. Bulletin, Ministry Natural Resour., Beijing, China, 2023.

[3] O. Korup and A. Stolle, "Landslide prediction from machine learning," *Geol. Today*, vol. 30, no. 1, pp. 26–33, Jan. 2014.

[4] R. S. Hadi, K. H. Ghazali, Ir. Z. Khalidin, and M. Zeehaida, "Human parasitic worm detection using image processing technique," in *Proc. Int. Symp. Comput. Appl. Ind. Electron. (ISCAIE)*, Kuala Lumpur, Malaysia, Dec. 2012, pp. 196–201.

[5] B.-G. Chae, H.-J. Park, F. Catani, A. Simoni, and M. Berti, "Landslide prediction, monitoring and early warning: A concise review of state-of-the-art," *Geosci. J.*, vol. 21, no. 6, pp. 1033–1070, Dec. 2017.

[6] F. S. Tehrani, M. Calvello, Z. Liu, L. Zhang, and S. Lacasse, "Machine learning and landslide studies: Recent advances and applications," *Natural Hazards*, vol. 114, no. 2, pp. 1197–1245, Nov. 2022.

[7] S. D. Pardeshi, S. E. Autade, and S. S. Pardeshi, "Landslide hazard assessment: Recent trends and techniques," *SpringerPlus*, vol. 2, no. 1, p. 523, Oct. 2013.

[8] F. Huang, J. Huang, S. Jiang, and C. Zhou, "Landslide displacement prediction based on multivariate chaotic model and extreme learning machine," *Eng. Geol.*, vol. 218, pp. 173–186, Feb. 2017.

[9] Y. Meng, Y. Qin, Z. Cai, B. Tian, C. Yuan, X. Zhang, and Q. Zuo, "Dynamic forecast model for landslide displacement with step-like deformation by applying GRU with EMD and error correction," *Bull. Eng. Geol. Environ.*, vol. 82, no. 6, p. 211, Jun. 2023.

[10] T. Wang, M. Zhang, Q. Yu, and H. Zhang, "Comparing the applications of EMD and EEMD on time–frequency analysis of seismic signal," *J. Appl. Geophys.*, vol. 83, pp. 29–34, Aug. 2012.

[11] Z. H. Wang, W. Nie, and H. H. Xu, "Prediction of landslide displacement based on EEMD-prophet-LSTM," *J. Univ. Chin. Acad. Sci.*, vol. 40, no. 4, pp. 514–522, Apr. 2023.

[12] L. W. Li, Y. P. Wu, and F. S. Miao, "Research on landslide displacement prediction based on variational mode decomposition and GWO-MIC-SVR model," *Rock Mech. Eng.*, vol. 37, no. 6, pp. 1395–1406, Jun. 2018.

[13] W. Chu, G. Y. Wang, and T. Liu, "Rolling bearing fault diagnosis based on VMD combined with K-SVD with sparrow algorithm parameters optimization," *Noise Vib. Control*, vol. 42, no. 4, Apr. 2022, Art. no. 100.

[14] Y. H. Jiang, W. Wang, and L. F. Zou, "Research on dynamic prediction model of landslide displacement based on particle swarm-variational modal decomposition, nonlinear autoregressive neural network and gated cyclic unit," *Rock Soil Mech.*, vol. 43, no. 1, pp. 601–612, Jan. 2022.

[15] W. Yuan, R. F. Sun, H. Y. Zhong, and M. Jaboyedoff, "Research on comprehensive deformation prediction and monitoring and early warning methods of step-type landslides," *J. Hydraul. Eng.*, vol. 54, no. 4, pp. 461–473, Apr. 2023.

[16] C. H. Wang, Y. J. Zhao, and W. Guo, "Landslide displacement EEMD-SVR prediction model," *J. Surveying Mapping*, vol. 51, no. 10, pp. 2196–2204, Oct. 2022.

[17] F. M. Huang, Z. Ye, and C. Yao, "Uncertainty in landslide susceptibility prediction: The impact of different attribute interval divisions of environmental factors and different data-driven models," *Earth Sci.*, vol. 45, no. 12, pp. 4535–4549, Dec. 2020.

[18] Q. Liu and W. Yi, "Landslide displacement prediction based on inducing factor response and BP neural network," *J. Chin. Three Gorges Univ.*, vol. 41, no. 3, pp. 41–45, Mar. 2019.

[19] K. Liao, Y. Wu, and L. Li, "Landslide displacement prediction based on time series and GWO-ELM model," *J. Cent. South Univ.*, vol. 50, no. 3, pp. 619–626, Mar. 2019.

[20] G. R. Pan, "Research on dynamic deformation prediction model based on time series analysis," *J. Wuhan Univ.*, vol. 30, no. 6, pp. 483–487, Jun. 2005.

[21] L. Yang, Y. Q. Wu, and J. L. Wang, "A review of recurrent neural network research," *J. Comput. Appl.*, vol. 38, no. 2, pp. 1–6, 2018.

[22] F. Wang, "Research progress and application of LSTM recurrent neural network," Doctoral dissertation, Heilongjiang Univ., Harbin, China, 2021.

[23] Z. Lin, X. Sun, and Y. Ji, "Landslide displacement prediction model using time series analysis method and modified LSTM model," *Electronics*, vol. 11, no. 10, p. 1519, May 2022.

[24] S. Zhang, H. Jia, C. Wang, X. Wang, S. He, and P. Jiang, "Deep-learning-based landslide early warning method for loose deposits slope coupled with groundwater and rainfall monitoring," *Comput. Geotechnics*, vol. 165, Jan. 2024, Art. no. 105924.

[25] X. J. Liu, X. S. Liu, and L. W. Zhang, "Dam deformation prediction based on VMD-HPO-BiLSTM," *J. Geodesy Geodynamics*, vol. 43, no. 8, pp. 851–855, 2023.

[26] H. Y. Liu, W. T. Chen, Y. Y. Li, Z. Y. Xu, and X. J. Li, "Application of a new comprehensive model based on EEMD-CNN-LSTM in landslide displacement prediction," *Geomechanics*, vol. 30, no. 4, pp. 633–646, 2024.

[27] D. Zhang, J. Yang, F. Li, S. Han, L. Qin, and Q. Li, "Landslide risk prediction model using an attention-based temporal convolutional network connected to a recurrent neural network," *IEEE Access*, vol. 10, pp. 37635–37645, 2022.

[28] J. Xue and B. Shen, "A novel swarm intelligence optimization approach: Sparrow search algorithm," *Syst. Sci. Control Eng.*, vol. 8, no. 1, pp. 22–34, Jan. 2020.

[29] C. Li, B. Feng, S. Li, J. Kurths, and G. Chen, "Dynamic analysis of digital chaotic maps via state-mapping networks," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 66, no. 6, pp. 2322–2335, Jun. 2019.

[30] G. Long, X. Chai, Z. Gan, D. Jiang, X. He, and M. Sun, "Exploiting one-dimensional exponential Chebyshev chaotic map and matching embedding for visually meaningful image encryption," *Chaos, Solitons Fractals*, vol. 176, Nov. 2023, Art. no. 114111.

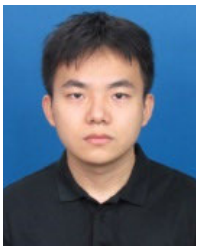
[31] P. Barthelemy, J. Bertolotti, and D. S. Wiersma, "A Lévy flight for light," *Nature*, vol. 453, no. 7194, pp. 495–498, 2008.

[32] K. Dragomiretskiy and D. Zosso, "Variational mode decomposition," *IEEE Trans. Signal Process.*, vol. 62, no. 3, pp. 531–544, Feb. 2014.

[33] Y. Xu, Y. F. Xiang, and T. X. Ma, "VMD-GRU short-term electric load forecasting model based on particle swarm algorithm optimized parameters," *J. North China Electric Power Univ.*, vol. 50, no. 1, pp. 38–47, 2023.

[34] Y. Yang, H. Liu, L. Han, and P. Gao, "A feature extraction method using VMD and improved envelope spectrum entropy for rolling bearing fault diagnosis," *IEEE Sensors J.*, vol. 23, no. 4, pp. 3848–3858, Feb. 2023.

[35] G. Duan, Y. Su, and J. Fu, "Landslide displacement prediction based on multivariate LSTM model," *Int. J. Environ. Res. Public Health*, vol. 20, no. 2, p. 1167, Jan. 2023.
[36] Y. Wang, S. Li, and B. Li, "Deformation prediction of cihaxia landslide using InSAR and deep learning," *Water*, vol. 14, no. 24, p. 3990, Dec. 2022.
[37] C. Zhang, Y. Guo, and M. Li, "A review on the development and application of artificial neural network models," *Comput. Eng. Appl.*, vol. 57, no. 11, pp. 57–69, 2021.
[38] P. Xie, A. Zhou, and B. Chai, "The application of long short-term memory(LSTM) method on displacement prediction of multifactor-induced landslides," *IEEE Access*, vol. 7, pp. 54305–54311, 2019.
[39] A. Sherstinsky, "Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network," *Phys. D, Nonlinear Phenomena*, vol. 404, Mar. 2020, Art. no. 132306.
[40] S. Siami-Namini, N. Tavakoli, and A. S. Namin, "The performance of LSTM and BiLSTM in forecasting time series," in *Proc. IEEE Int. Conf. Big Data (Big Data)*, Los Angeles, CA, USA, Dec. 2019, pp. 3285–3292.
[41] S. Siami-Namini, N. Tavakoli, and A. S. Namin, "A comparative analysis of forecasting financial time series using ARIMA, LSTM, and BiLSTM," 2019, *arXiv:1911.09512*.
[42] T. J. Ham, S. J. Jung, S. Kim, Y. H. Oh, Y. Park, Y. Song, J.-H. Park, S. Lee, K. Park, J. W. Lee, and D.-K. Jeong, "A^3: Accelerating attention mechanisms in neural networks with approximation," in *Proc. IEEE Int. Symp. High Perform. Comput. Archit. (HPCA)*, San Diego, CA, USA, Feb. 2020, pp. 328–341.
[43] J. Liu, J. Liu, and X. Luo, "Research progress in attention mechanism in deep learning," *Chin. J. Eng.*, vol. 43, no. 11, pp. 1499–1511, 2021.
[44] Y. Sun, S. Liu, and L. Li, "Grey correlation analysis of transportation carbon emissions under the background of carbon peak and carbon neutrality," *Energies*, vol. 15, no. 9, p. 3064, Apr. 2022.
[45] Y. Xing, J. Yue, and C. Chen, "Interval estimation of landslide displacement prediction based on time series decomposition and long short-term memory network," *IEEE Access*, vol. 8, pp. 3187–3196, 2020.
[46] Y. Q. Sun, D. Y. Li, K. L. Yin, and L. X. Chen, "Prediction of intermittent landslide activity in the accumulation layer of the three Gorges reservoir area: A case study of the Baishuihe landslide," *Geological Sci. Technol. Bull.*, vol. 38, no. 5, p. 195, 2019.
[47] R. Chen, X. G. Fan, and Y. P. Wu, "Analysis of multi-field information correlation criteria of Baishuihe landslide based on data mining technology," *Chin. J. Geological Hazard Control*, vol. 32, no. 6, pp. 1–8, 2021.
[48] W. Yi. (2016). *Basic characteristics and monitoring data of Baishui River landslide in Zigui County, Three Gorges Reservoir Area of the Yangtze River From 2007 to 2012*. National Data Center for Glacier, Frozen Soils and Deserts. [Online]. Available: http://www.ncdc.ac.cn
[49] M. Shang, F. Liao, and R. Ma, "Quantitative analysis of the correlation between Baijiabao landslide deformation, reservoir water level and rainfall," *J. Eng. Geol.*, vol. 29, no. 3, pp. 742–750, 2021.
[50] J. F. Tang, X. M. Tang, and P. Xiao, "Analysis of seepage stability of large landslide under the action of reservoir water level fluctuation and rainfall," *Geological Sci. Technol. Bull.*, vol. 40, no. 4, pp. 153–161, 2021.
[51] S. Mirjalili, S. M. Mirjalili, and A. Lewis, "Grey wolf optimizer," *Adv. Eng. Softw.*, vol. 69, pp. 46–61, Jan. 2014.
[52] S. Mohapatra and P. Mohapatra, "Fast random opposition-based learning golden jackal optimization algorithm," *Knowl.-Based Syst.*, vol. 275, Sep. 2023, Art. no. 110679.



SHUAI REN was born in Xi'an, Shaanxi, in 2000. He received the master's degree from Guilin University of Electronic Technology, in 2024. He is currently pursuing the Ph.D. degree with Universiti Malaysia Pahang Al-Sultan Abdullah, Malaysia. His main research interests include disaster warning and machine learning.

KAMARUL HAWARI GHAZALI (Senior Member, IEEE) is currently a Professor with the Faculty of Electrical and Electronics Engineering Technology, Universiti Malaysia Pahang Al-Sultan Abdullah (UMPSA). He specializes in artificial intelligence, machine learning, and image processing, with applications in signal analysis and smart manufacturing. He has led international research collaborations and delivered keynotes at global conferences. His research interests include deep learning for predictive systems, smart prosthetic technologies, and sustainable energy solutions. He is a Senior Member of Chartered Engineer.

YUANFA JI was born in Shandong, China, in 1975. He received the M.S. degree from Guilin University of Electronic Technology, in 2004, and the Ph.D. degree in science and engineering from National Astronomical Observatories, Chinese Academy of Sciences, in 2008. From 2009 to 2011, he was an Assistant Professor with Guilin University of Electronic Technology, where he has been a Professor with the School of Information and Communication, since 2012. He is the author of one book, more than 100 articles, and more than 30 inventions. His research interests include satellite communications, satellite navigation, real-time kinematic positioning, and navigation receivers.

SAMRA UROOJ KHAN received the B.S. degree from Pakistan and the M.S. degree in electrical engineering from Universiti Tun Hussein Onn Malaysia (UTHM). She is currently pursuing the Ph.D. degree with Universiti Malaysia Pahang Al-Sultan Abdullah, Pakistan. Her research interests include wireless communications, image processing, machine learning, and signal processing.