

Research Article

Landslide Displacement Prediction Based on Transfer Learning and Bi-GRU

Haiqing Zheng ¹, Mengfan Duan,¹ Xiaoyun Sun ¹, Guang Han,¹ and Qiang Jin²

¹School of Electrical and Electronic Engineering, Shijiazhuang Tiedao University, Shijiazhuang, Hebei 050043, China

²Hebei Jinyu Zenith Cement Co., Ltd, Shijiazhuang, Hebei 050223, China

Correspondence should be addressed to Xiaoyun Sun; sunxy1971@126.com

Received 6 August 2021; Revised 31 August 2022; Accepted 22 September 2022; Published 11 October 2022

Academic Editor: Kathiravan Srinivasan

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

Predicting slope deformation prediction is crucial for early warning of slope failure, preventing damage to properties, and saving human lives. However, in practice, equipment maintenance causes discontinuity in the displacement data, and the traditional prediction models based on deep networks do not perform well in this case. To solve the problem of prediction accuracy in case of discontinuous and inadequate data, we propose a combined displacement prediction model that integrates the bidirectional gated recurrent unit (Bi-GRU), attention mechanism, and transfer learning. The Bi-GRU is employed to extract the forward and backward characteristics of displacement series, and the attention mechanism is utilized to give different weights to the extracted information so as to highlight the critical information. Transfer learning is used to guarantee prediction accuracy in case of discontinuous and limited data. The model is then employed to predict the slope displacement of the JinYu Cement Plant in China. Finally, the modeling results excellently agree with measured displacement, especially in case of insufficient sample data.

1. Introduction

There have been enormous casualties and economic losses due to landslides recently. On July 23, 2019, a landslide occurred in Shuicheng, China, which affected more than 1,600 people and caused an economic loss of 190 million RMB. On July 16, 2020, three landslides in Dunhao, China, resulted in two deaths and four missing people. Due to considerable damage from landslide disasters, it is necessary to establish reliable early warning models [1]. According to Web GIS techniques, areas at risk from landslides have been estimated by using various geospatial data [2]. Landslides are generally believed to be governed by their physical characteristics and triggering factors. The physical characteristics primarily include stratigraphic lithology, pore-fluid pressure, and gravitational stress. The triggering factors are much more diverse. Rainfall, earthquakes, human activities, vegetation cover, water level fluctuations, and erosion processes may cause slope failure [3]. The variation of the slope displacement can reflect the instability of landslides and has

been considered a critical index for evaluating the stability and status of landslides [4]. The monitored series of slope displacements reflect the nonlinear dynamic evolution of the slope with the dual impact of internal and external environments.

By analyzing the monitored series of slope displacement, various displacement prediction models are constructed based on intelligent algorithms, such as the gray model, the support vector regression model, the wavelet neural network, and the particle swarm optimization (PSO). Combined with actual slope engineering, the mechanism and law of slope deformation are grasped to provide guidance for landslide prevention.

Kuradusenge et al. [5] proposed random forest (RF) and logistic regression (LR) models to predict rainfall-induced landslides in Rwanda. They used rainfall data sets and various internal and external parameters for landslide prediction, thereby improving its accuracy. In another work, Yan et al. [6] employed the Holt-Winters damped model to predict the deep horizontal displacement of sloping soil. He

et al. [7] presented a machine-learning-aided stochastic reliability analysis of spatially variable slopes, and it significantly reduced the computational efforts and provided a complete statistical description of the safety factor with reasonable accuracy. Wang et al. [8] also developed a direct interval prediction approach based on least-squares support-vector machines (LS-SVM) and the differential search algorithm to predict the landslide displacements in the Three Gorges Reservoir area. In another work, Ma et al. [9] proposed measures based on mutual information (MI) for input variable selection (IVS) and optimized support vector regression (SVR) for predicting the displacement of seepage-driven landslides. Their experimental results demonstrated that MI-based measures could effectively identify relevant or critical variables, and the optimized SVR methods based on variable-reduced inputs could significantly enhance prediction performance.

Further, Wang et al. [10] compared the performance of several popular machine learning methods in predicting the displacement of reservoir landslides. Ray et al. [11] also developed an artificial neural network (ANN) to evaluate the safety factor of Shiwalik Hills in the Himalayan region. Shihabudheen et al. [12] developed an extreme learning, adaptive, neuro-fuzzy inference system (ELANFIS) using the empirical-mode decomposition (EMD) technique to predict the displacement of step-like landslides.

However, the prediction accuracy of the above methods depends on their hyper-parameters, which usually differ for each time series. Furthermore, misspecification reduces prediction accuracy by generating overfitted or underfitted models [13]. Table 1 lists the characteristics of some models for predicting slope deformation.

Since frameworks based on deep learning perform well in terms of discovering and extracting the internal structure of data, researchers are inspired to employ deep architectures in preference prediction tasks. Ref. [14] proposed an algorithm based on the deep recurrent neural network (RNN) called DeepVM to predict vehicle mobility in a future period of several (or tens) of minutes. Ref. [15] also developed a deep neural network named convolution-based LSTM network by mining the in situ vibration data to predict the remaining useful life of rotating machinery. In another work [16], a deep-learning-based approach was developed to characterize the probability density function (PDF) of wind for the following hours, where the convolutional neural network (CNN) and the gated recurrent unit (GRU) were employed to capture spatial and temporal features simultaneously. A composite GRU-Prophet model was also constructed to predict sales volume [17].

Further, Ref. [18] utilized the bidirectional long short-term memory (Bi-LSTM) and bidirectional GRU (Bi-GRU) to predict the network traffic matrix, and the results showed that the bidirectional models offered the ability to look into data series in two opposite directions, allowing the acquisition of the full knowledge of past and future statuses for better prediction results. The social force model (SFM) was incorporated into an original LSTM network to predict vessel trajectories, taking account of the social force-driven LSTM network and the mixed loss function in artificial intelligence- (AI-) powered maritime Internet of Things (IoT)

systems to achieve robust prediction [19]. Ref. [20] proposed a Bi-GRU-based framework to predict a rating for customer review, and the experimental results demonstrated that the developed framework could significantly enhance the rating prediction in the case of balanced and imbalanced data sets. The LSTM, GRU, RNN, Bi-GRU, and Bi-LSTM were also developed to forecast solar irradiance, and the experimental results showed that the Bi-GRU model performed better than the others in terms of the training time, the trainable parameters, and the epoch ratio [21]. It can be stated that, in a given time, bidirectional models are more reliable due to their capacity to learn more abstract features of time-series data sets.

The remainder of the paper is organized as follows. Section 2 reviews the related literature on slope displacement prediction, and Section 3 presents the architecture of the proposed displacement prediction model and its execution procedure. Section 4 reports the experimental results of a case study in the JinYu Cement Plant, an overview of the study site, the prediction results, and the model performance evaluation. Section 5 summarizes the model outcomes and limitations.

2. Related Works

The related literature has addressed slope displacement prediction from different perspectives [22–26]. Here, we review the recent progress in displacement prediction methods based on deep learning.

Ref. [27] proposed a dynamic model based on the LSTM neural network to predict landslide displacement. The accumulated displacement was decomposed into a trending and periodic term, where a cubic polynomial function was employed to estimate the trend displacement, and an LSTM model was used to predict the periodic displacement. The performance of the model was validated with the observations of two step-wise landslides in the Three Gorges Reservoir area, the Baishuihe and Bazimen landslides. Another work introduced two new concepts, namely, the trend sequence and the sensitivity states, to quantitatively characterize the response of landslide displacement to external factors and the internal states of landslides, respectively [28]. Then, the PSO algorithm and the SVR method were used to obtain the trend sequence, and the LSTM neural network was employed to predict the sensitivity states. Taking the Baishuihe landslide located in the Three Gorges Reservoir area as a case study, the proposed model showed satisfactory performance. Another study [29] investigated a coupling prediction model based on the double moving average (DMA) method and the LSTM network to improve the accuracy of landslide displacement point prediction and quantify the uncertainties associated with the predicted values.

Moreover, three prediction models, namely the LSTM, GRU, and RF, were utilized to estimate the periodic and total accumulated displacement of three step-wise landslides in the Three Gorges Dam Reservoir area [30]. Zhang et al. [31] also developed a novel and dynamic model to predict the displacements of step-like landslides. To this end, variational mode decomposition (VMD) was used to decompose the cumulative displacement into stochastic, periodic, and

TABLE 1: The characteristics of some models of predicting slope deformation.

Category	Commonly used model	Input data or characteristics	Limitations
Statistics-based model	Auto-regressive moving average model (ARMA)	The correlation analysis of the input series is required, and the order and parameters of the model should be determined in advance.	Since a variety of factors affect landslide deformation, the statistical model cannot reflect the fluctuation in the landslide displacement with the influencing factors.
	Gray model and related improved models	The original time series should be accumulated to yield the once accumulating generation operator (1-AGO) data sequence.	
Machine-learning-based model	Extreme learning machine model	The inputs are usually the selected inducing factors or features.	Static fitting is performed on historical data, which cannot reflect the nature of the dynamic evolution of landslides
	Support-vector machine (SVM) model		
	Neural network model	The input can be original displacement series and various influencing factors.	When the sample data are limited, overfitting can easily occur, resulting in poor prediction accuracy.
Deep-learning-based models, such as long short-term memory (LSTM) and gated recurrent unit (GRU) models			

trend components. A polynomial expression with an optimized order fitted the trend displacement, and the bidirectional long short-term memory model dynamically modeled the periodic and stochastic displacements. The experiments demonstrated that the proposed model better predicted the displacement of step-like landslides. In another work [32], a novel prediction model based on the graph convolutional network was derived to estimate the slope deformation; the model considered the spatial correlations between all points in the entire displacement monitoring system. Furthermore, the complete-ensemble empirical-mode decomposition with adaptive noise (CEEMDAN) algorithm was employed to divide the total displacement into a trending, periodic, and residual term [33]. Then, a novel model based on the LSTM was established to predict the displacement of landslides.

However, in real-world scenarios, the training data sets for slope displacement are insufficient, which somewhat restricts the applications of the above methods. In addition, monitoring equipment can be destroyed under complex external conditions, and monitoring interruption results in data missing for model training. The transfer learning (TL) method solves the above problems in such cases because the knowledge acquired from the source tasks with sufficient training data can be applied to the target task without enough training data. Therefore, this paper combines the Bi-GRU network, attention mechanism, and transfer learning to construct a model for predicting the slope displacement of the landslide in Hebei JinYu Zenith Cement Co., Ltd., China. Concurrently, some ordinary networks such as SVR, LSTM, GRU, Bi-LSTM, Bi-GRU, and Bi-GRU with attention mechanism and the transfer network proposed in this work are invoked for comparison.

3. Establishment of the Slope Displacement Prediction Model

Predicting slope displacement can be regarded as a time-series prediction problem, and the Bi-GRU can deeply capture the relationship between the forward and backward

characteristics of the displacement data. Bidirectionally handling data in the learning process offers the opportunity to have a strong representation of the features to better predict the future step. The attention mechanism can pay attention to only relevant pieces of information, and specifically, the importance of pieces of information is determined by their probability distribution, which eliminates unnecessary information.

During the process of slope deformation, the distribution of displacement monitoring data changes remarkably over time, and the maintenance of monitoring equipment causes discontinuity in the data, resulting in the discontinuity between the historical and newly collected data on displacement. Discontinuity in the training data leads to the deterioration of network prediction performance, so the transfer learning strategy is introduced to solve this problem. The model trained with historical data is used as the source domain model, and the knowledge and skills learned from the source domain model are applied to the new data (the target task) over a long period to assist in predicting new data. In this way, knowledge acquisition is no longer from scratch, which can enhance prediction accuracy to a certain extent.

Figure 1 shows the technical method this paper employs to predict the slope displacement accurately. First, the collected displacement data are divided into the source and target data. Afterward, a displacement prediction model based on the Bi-GRU and the attention mechanism (denoted as Bi-GRU-ATT) is constructed, and model training is conducted using the source data. Then, a small amount of target data is used to fine-tune the prediction model so as to obtain the target model. After that, the prediction performance of the model is evaluated using the testing data set. Finally, the prediction results of the model are denormalized to estimate the displacement of the landslide.

Figure 2 illustrates the model structure diagram, comprising the input layer, the Bi-GRU layer, the attention layer, the fully connected layer, and the output layer.

3.1. Bidirectional Gated Recurrent Unit. The recurrent neural network focuses on time-series problems because of its

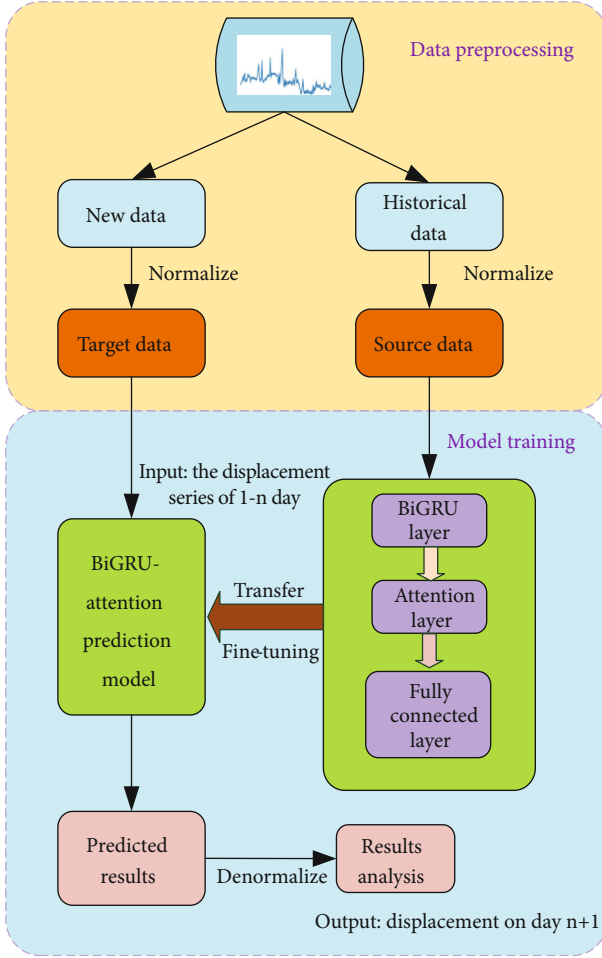


FIGURE 1: Technical route of slope displacement prediction.

advantages over sequence dependence. However, due to gradient exploding and vanishing, there is a length limitation when applying the RNN algorithm [34]. Subsequently, variants of RNN, such as the LSTM and GRU neural network, are proposed to deal with long-sequence prediction problems. The GRU retains some properties of the LSTM, and simplifies the structure, which can realize information forgetting and memorizing with the same gate, thereby giving rise to fewer parameters and faster convergence [35]. Figure 3 depicts the structure of the gated recurrent unit.

The GRU model consists of two gates, reset gate and update gate. The update gate controls the previous information that will be carried over to the current layer, while the reset gate decides the amount of information to forget. As seen in Figure 3, the GRU transmits the input state x_t at the current moment and the hidden layer's output state at the previous moment to the reset gate, which determines how much memory information is saved. It can be formulated as

$$r_t = \sigma(W^r x_t + U^r h_{t-1} + b^r), \quad (1)$$

where r_t is the output of reset gate, σ is the logistic sigmoid function, x_t and h_{t-1} are the input and the previous hidden

state, respectively. W^r and U^r are learned weight matrices. b^r is the bias.

Similarly, the output of update gate is computed by

$$z_t = \sigma(W^z x_t + U^z h_{t-1} + b^z). \quad (2)$$

The new memory state \bar{h}_t can be computed by

$$\bar{h}_t = \tanh\left(W^h x_t + r_t \circ \left(U^h h_{t-1} + b^h\right)\right). \quad (3)$$

The final hidden state h_t can be computed by

$$h_t = z_t \circ h_{t-1} + (1 - z_t) \circ \bar{h}_t. \quad (4)$$

where \circ represents the Hadamard product. The gating signal z_t ranges from 0 to 1. The closer the gating signal is to 1, the more data will be memorized, whereas the closer to 0, the more forgotten.

The unidirectional GRU model uses previous information to predict follow-up information, while the bidirectional GRU can comprehensively learn the time-correlated information in both forward and backward directions simultaneously to improve prediction accuracy [36]. The Bi-GRU extends the unidirectional GRU by introducing a second layer, as shown in Figure 4.

The horizontal direction calculates the forward GRU hidden vector (\vec{h}_t) and the backward GRU hidden vector (\overleftarrow{h}_t) at each time step (t) simultaneously. The vertical direction represents a unidirectional flow from the input layer to the hidden layer and then to the output layer. Connecting the two hidden states can calculate the final prediction of the Bi-GRU, as described below:

$$\vec{h}_t = GRU\left(x_t, \vec{h}_{t-1}\right), \quad (5)$$

$$\overleftarrow{h}_t = GRU\left(x_t, \overleftarrow{h}_{t-1}\right), \quad (6)$$

$$y_t = W_t \vec{h}_t + V_t \overleftarrow{h}_t + b_y, \quad (7)$$

where $GRU(\cdot)$ represents GRU function, W_t, V_t are the weights of forward GRU and backward GRU, respectively, and b_y is the bias of the output layer.

Predicting the displacement of a slope can be considered a regression problem related to time series. The bidirectional structure offers the ability to look into the displacement series in two opposite directions, which allows acquiring the full knowledge of past and future statuses for better prediction results.

3.2. Attention Mechanism. Generally, the GRU model can effectively avoid the phenomenon of gradient disappearance, however, when the time series is too long, the model may experience the phenomenon of gradient disappearance. Therefore, an attention mechanism is introduced to reduce the loss of historical information and strengthen the influence

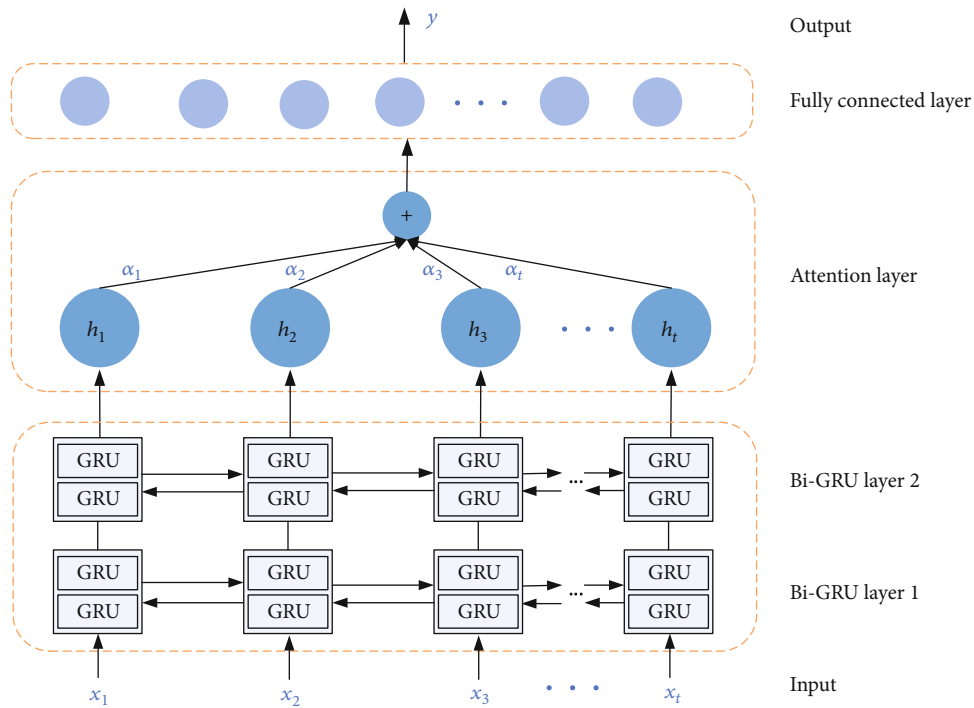


FIGURE 2: The model structure diagram.

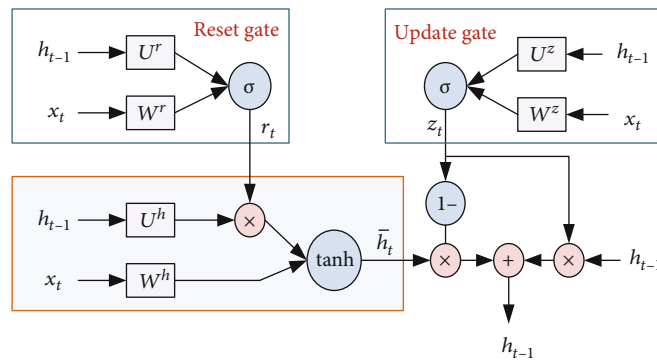


FIGURE 3: The structure of GRU.

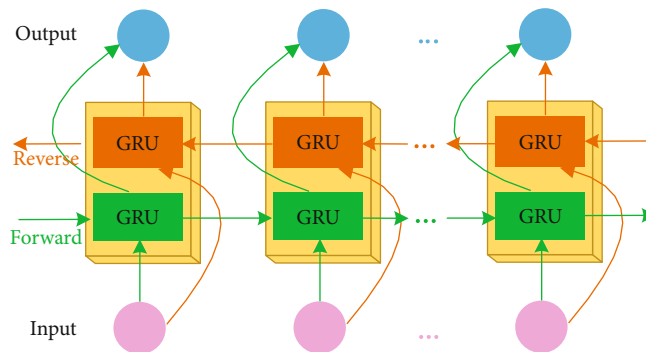


FIGURE 4: The structure of Bi-GRU.

of important information, which is one of the most influential ideas in the field of deep learning. The attention mechanism assigns different weights to the Bi-GRU hidden layer, which highlights the critical information, thereby optimizing the out-

put value and improving the prediction accuracy. Figure 5 depicts a schematic of the attention mechanism.

The attention mechanism can simply be a weighted summation [37]. It starts by calculating the importance of the

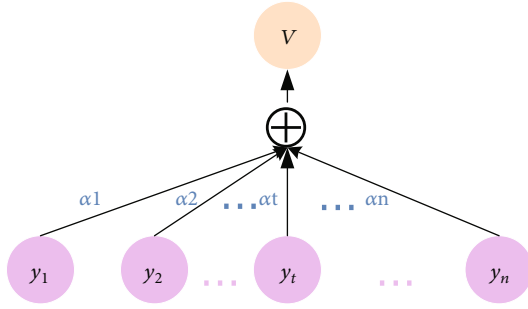


FIGURE 5: The schematic of the attention mechanism.

input information; then, the softmax function is utilized to ensure that the sum of all the weights equals 1.0. The weights and the input information are multiplied correspondingly, and are eventually summed to obtain the output. The relevant equations are as follows:

$$\alpha_t = \text{softmax}(u_t) = \frac{\exp(u_t)}{\sum_{t=1}^n \exp(u_t)}, \quad (8)$$

$$u_t = \tanh(W_d y_t + b_d), \quad (9)$$

$$V = \sum_{t=1}^n \alpha_t u_t, \quad (10)$$

where n is the total time steps of the input sequence, α_t is the weight calculated for each state u_t at each time step, W_d is the weight matrix, b_d is the weight offset, and y_t is the hidden state of Bi-GRU.

3.3. Transfer Learning. Transfer learning is a primary method to solve the problem of insufficient available label data in machine learning [38]. This method relaxes the assumption that the training and testing data sets are identically distributed so as to realize the knowledge transfer from the source domain to the target domain. After establishing the training network based on the Bi-GRU and the attention mechanism, this work uses the source data set to train the training network so as to obtain the source model. Then, the target data set is employed to fine-tune the full connection layer of the source model so as to obtain the target model for predicting small data sets. Figure 6 shows the specific process.

Deep learning prediction models are prone to overfitting for small data sets, and their prediction accuracy is not high [21, 39]. On the other hand, obtaining a perfect data set is expensive and time-consuming. Moreover, the collected data on slope displacement are usually insufficient, and the data reflecting the slope inclining to slide are limited, which affects the accuracy of early warning. Transfer learning is essential to solve the fundamental problem of insufficient training data. Therefore, concerning the less effective sample data during a long period of slope deformation, we introduce a transfer learning algorithm to solve the slope displacement prediction under limited sample data.

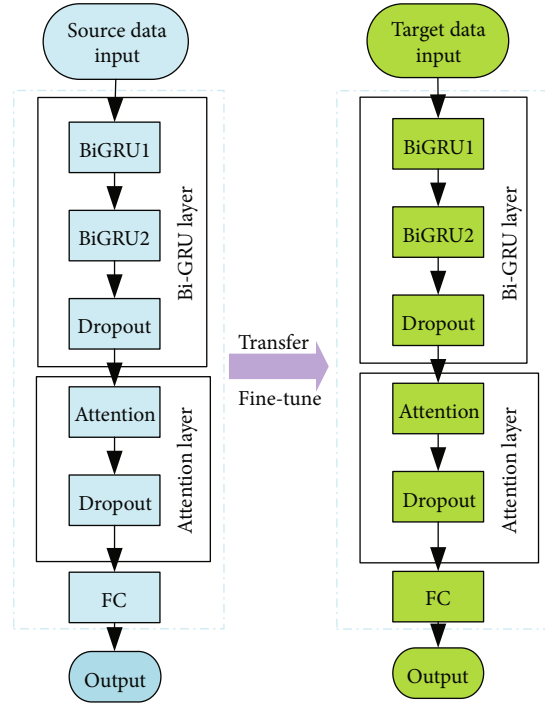


FIGURE 6: The specific process of transfer learning.

4. Case Study

4.1. Project Description. The case study area is the slope of the JinYu Cement Plant in the Luquan District, Shijiazhuang, China. The JinYu Cement Plant established a highly integrated global navigation satellite system (GNSS) online monitoring system, realizing the comprehensive network monitoring of the slope with high efficiency, sensitivity, quick response, and reliable operation, to improve the slope monitoring and the emergency response level. Figure 7 shows the location and actual view of the slope studied, and Figure 8 depicts the layout of the monitoring points and installed GNSS monitoring equipment.

Nine monitoring points are arranged on this slope, as shown in Figure 8. Table 2 presents some of the data collected by monitoring point G112 in 2019, where X, Y, and Z represent the displacement in the north, east, and vertical directions of the coordinate system, respectively. A negative value indicates the opposite direction.

4.2. Data Preprocessing. The experiment selects 798 displacement data obtained from monitoring point G112 from 2019 to 2021 as the experimental objects, and the first 648 data from June 2019 to January 2021 are used as the source data. Due to the equipment overhaul, 150 groups of data from March 2021 to July 2021 are employed as the target data. For target model training, the first 110 data from the target data set are selected as the training data set. After fine-tuning the training model, the last 40 data are selected for its testing.



FIGURE 7: The location and actual view of the slope studied.



FIGURE 8: The layout of the monitoring points and installed GNSS monitoring equipment.

TABLE 2: Some of the data collected by monitoring point G112.

Date	X	Y	Z	Displacement
2019/6/1 17:09:50	-6.630	-3.980	-10.530	7.733
2019/6/2 17:10:02	-8.260	-4.350	-8.170	9.336
2019/6/3 17:13:28	-9.460	-4.530	-18.210	10.489
2019/6/4 17:16:37	-12.600	-3.570	-21.450	13.096
2019/6/5 17:17:00	-10.250	-1.930	-15.080	10.431
2019/6/6 13:27:51	-14.490	-2.390	-28.770	14.686
2019/6/7 17:01:55	-14.360	-6.550	-43.030	15.784
2019/6/8 17:09:17	-8.240	-4.840	-6.520	9.557
2019/6/9 17:08:36	-9.500	-5.720	-22.250	11.090
2019/6/10 17:16:25	-9.810	-3.520	-7.340	10.423
2019/6/11 17:16:58	-11.350	-2.070	-17.100	11.538
2019/6/12 17:15:23	-14.520	-5.010	-26.350	15.361
2019/6/13 17:15:26	-13.240	-3.490	-17.030	13.693
2019/6/14 17:16:55	-13.470	-1.800	-14.870	13.590

TABLE 3: Comparison of prediction accuracy between different prediction models.

Evaluation criteria	TRA-Bi-GRU-ATT	Bi-GRU-ATT	Bi-GRU	Bi-LSTM	GRU	LSTM	SVR
RMSE (best)	0.495	1.219	1.507	1.575	1.643	1.813	1.682
RMSE (mean)	0.543	1.318	1.527	1.597	1.699	1.838	1.682
MAE (best)	0.391	1.026	1.223	1.304	1.361	1.479	1.396
MAE (mean)	0.445	1.085	1.270	1.325	1.409	1.531	1.396

TABLE 4: Some prediction results of different models.

The measured displacement	Predictive value of displacement						
	TRA-Bi-GRU-ATT	Bi-GRU-ATT	Bi-GRU	Bi-LSTM	GRU	LSTM	SVR
7.181	7.276	7.551	7.676	7.686	7.749	7.695	7.694
6.987	7.301	7.447	7.622	7.637	7.634	7.711	7.640
7.655	7.721	7.750	7.805	7.811	7.813	7.953	7.863
8.807	8.570	8.246	8.232	8.213	8.293	8.186	8.149
10.066	9.322	8.969	8.921	8.899	8.793	8.586	8.570
10.542	9.495	9.220	9.113	9.017	9.033	8.627	8.729
4.307	5.526	6.244	6.401	6.459	6.486	7.189	6.645
3.529	4.230	5.858	6.059	6.197	6.385	6.686	6.385
3.798	4.412	6.076	6.375	6.398	6.490	6.751	6.475

To improve the training speed and prediction accuracy, we normalize the collected sample data as follows:

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

where x is the original data, and x_{\max} and x_{\min} indicate the maximum and the minimum in the sample, respectively.

The root-mean-square error (RMSE) and mean absolute error (MAE) are used for the evaluation criteria, which can be expressed by Equations (12) and (13), respectively.

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2}, \quad (12)$$

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

where y_i denotes the measured accumulative displacement of the slope, \hat{y}_i denotes the predictive accumulative displacement of the slope, and n represents the number of predictive values.

RMSE and MAE reflect the difference between the predicted value and the measured value and the dispersion degree of errors. The smaller the RMSE and MAE, the better the model performance.

4.3. Results and Discussion. According to the training process, as described in the previous section, the performance of the ordinary networks, i.e., the SVR, LSTM, GRU, Bi-

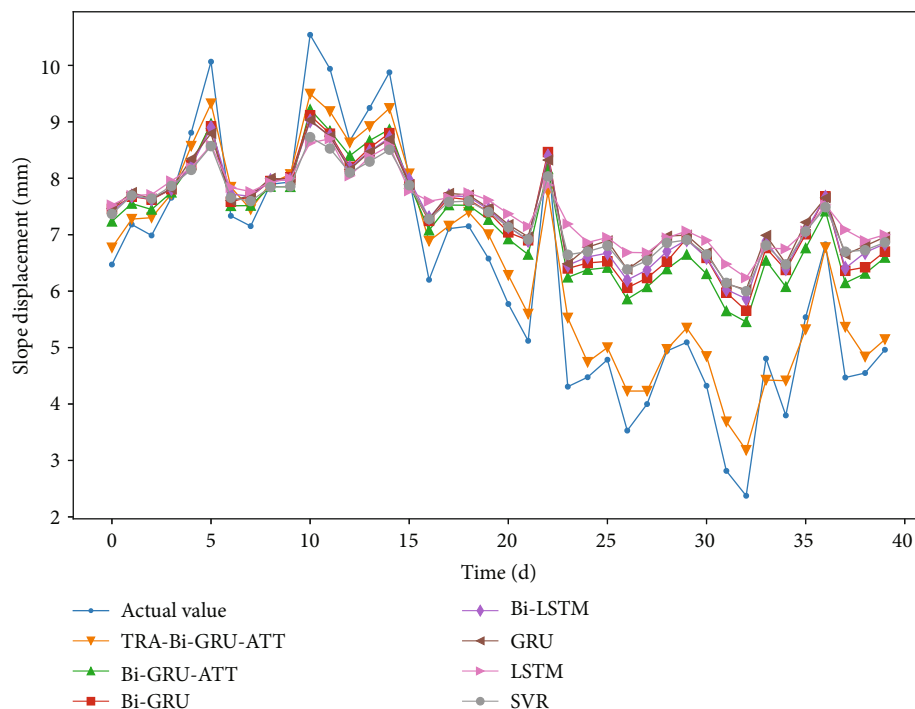


FIGURE 9: Comparison of the predictive results of monitoring point G112.

GRU, Bi-LSTM, and Bi-GRU with the attention mechanism, and the transfer network model, i.e., the Bi-GRU with the attention mechanism and transfer learning, is evaluated employing a testing data set.

Since deep learning methods have a specific degree of uncertainty, model training and predictions may differ each time. Each algorithm runs 100 times to evaluate the mean prediction accuracy of the model so as to reduce the uncertainty of the prediction results. For the SVR, the kernel function is linear and the penalty factor (C) is set at 200, the best prediction accuracy. The structural parameters and the training method of the Bi-GRU-ATT (the Bi-GRU model with the attention mechanism) and the TRA-Bi-GRU-ATT (the Bi-GRU model with the attention mechanism and transfer learning) are consistent: they comprise two layers of the Bi-GRU; the nodes in the hidden layers of the Bi-GRU-ATT and the TRA-Bi-GRU-ATT are also 64 and 128, respectively. The batch size and epochs are also set at 30 and 120, respectively.

Tables 3 and 4 and Figures 9–11 present the comparisons, and Figure 10 shows the absolute prediction errors of the different models. Since the stochastic gradient descent method maintains a special learning rate for training, the convergence speed is relatively low. Therefore, the Adam optimization algorithm is adopted to adapt each parameter to the learning rate.

The experimental results confirm that model TRA-Bi-GRU-ATT proposed in this paper achieves the highest degree of prediction accuracy. According to the best values of RMSE and MAE, the degree of single prediction accuracy can be higher in order: TRA – Bi – GRU – ATT > Bi – GRU – ATT > Bi – GRU > Bi – LSTM > GRU > SVR > LSTM. According to the mean values of the RMSE and the

MAE, the mean prediction accuracy can be higher in order: TRA – Bi – GRU – ATT > Bi – GRU – ATT > Bi – GRU > Bi – LSTM > SVR > GRU > LSTM. Although model TRA-Bi-GRU-ATT obtains the highest degrees of single and mean prediction accuracy, its stability is less than that of the SVR. During multiple experiments, when a suitable kernel function is chosen, the prediction accuracy of the SVR remains unchanged. However, due to too many parameters that must be tuned, the prediction stability of the TRA-Bi-GRU-ATT, Bi-GRU-ATT, Bi-GRU, Bi-LSTM, GRU, and LSTM is slightly poor.

Figure 11 shows the box plots comparing the prediction performance of the different models: the distribution of the absolute errors. The absolute error distribution of model TRA-Bi-GRU-ATT is more concentrated than that of the other models, and its prediction performance is superior to the others. Moreover, it can be concluded that transfer learning enhances the prediction performance of the model by extracting the critical features from the historical displacement sequence to assist with learning new data. It can entirely exploit the historical information within a long period and provide a good weight initialization for the training of the slope displacement model.

5. Conclusions and Future Perspectives

This work developed a predictive model combining transfer learning and the deep neural network, where the deep neural network adopted the Bi-GRU to extract the forward and backward information from the analyzed displacement series. The attention mechanism was introduced to highlight the critical information and reduce the loss of historical

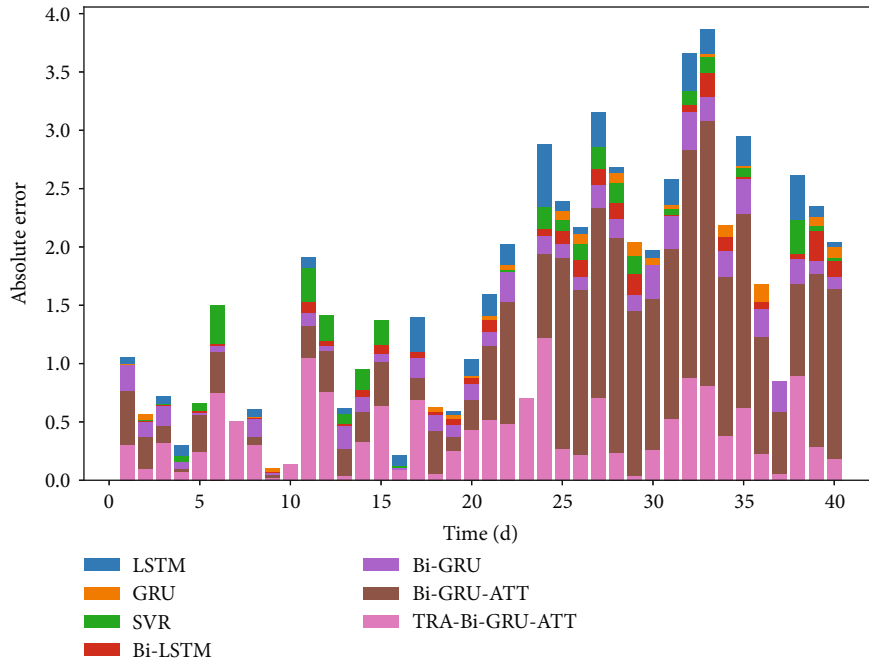


FIGURE 10: The absolute prediction errors of the different models.

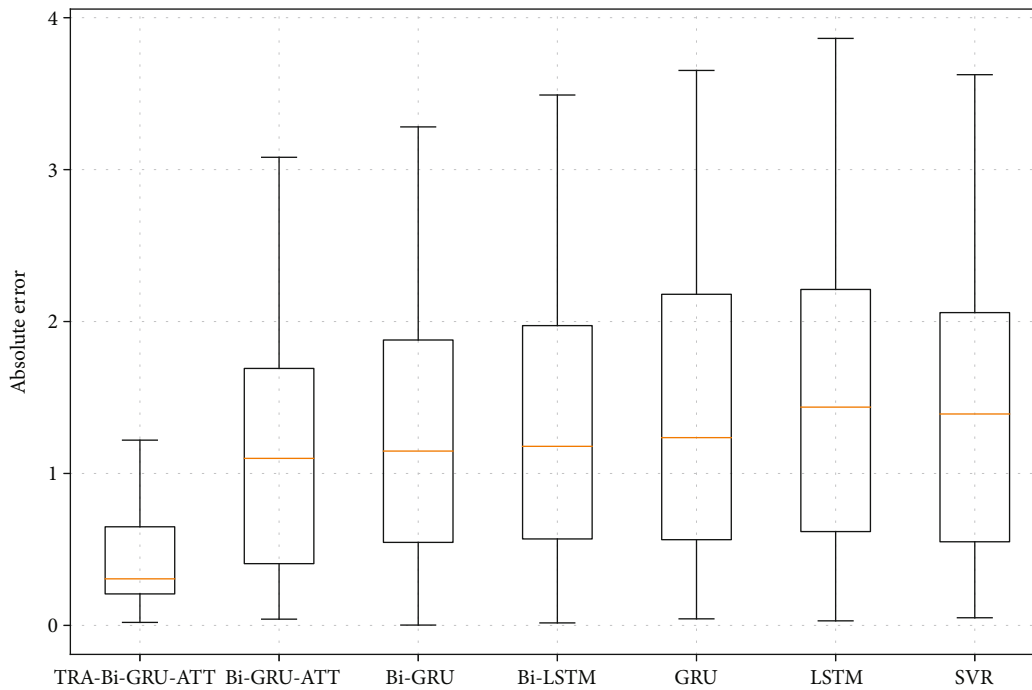


FIGURE 11: The box plots of absolute errors for different models.

information, thereby further stabilizing the model and enhancing its prediction efficiency.

Due to complex external conditions, monitoring systems can be destroyed, and monitoring interruption results in data missing for model training. Aiming at this problem, we introduce transfer learning employing historical data to assist with learning new data for displacement prediction. Taking the slope of the JinYu Cement Plant as an example,

the experimental results demonstrated that the model proposed herein offered a higher degree of accuracy in predicting the slope displacement and had the potential to estimate the displacement of landslides, particularly in case of limited sample data.

The shortcoming of the developed model was that the displacement prediction curve showed evolution characteristics from a single monitoring point. However, there were

nine monitoring points, and the spatial correlations between them were overlooked. Thus, it was difficult to reveal the displacement changes in the entire monitoring system. In the future, we plan to discuss the correlation between the displacement series of different monitoring points and devise a spatiotemporal prediction method based on time-series data from an entire monitoring system containing multiple points.

Furthermore, since machine learning methods have a specific degree of uncertainty, model training and predictions may differ. Ref. [10] compared the performance of five popular machine learning methods, and the experimental results showed that no method achieved the best results in all three aspects: the highest single accuracy, the mean accuracy, and the stability. Therefore, improving the stability of the prediction model while ensuring its accuracy can be one of the future research focuses.

Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Conflicts of Interest

The authors declare that there are no conflicts of interest.

Acknowledgments

This work was financed by the National Natural Science Foundation of China, under Grant No. 51674169, the Department of Education of Hebei Province of China under Grant No. ZD2019140, the Natural Science Foundation of Hebei Province of China under Grant No. F2019210243, and the S&T Program of Hebei under Grant No. 22375413D.

References

- [1] S. X. Huang, C. F. Li, and L. Luo, "Landslide monitoring network establishment within unified datum and stability analysis in the Three Gorges Reservoir area," *Journal of Sensors*, vol. 2017, Article ID 9819214, 13 pages, 2017.
- [2] N. Tengtrairat, W. L. Woo, P. Parathai, C. Aryupong, P. Jitsangiam, and D. Rinchumphu, "Automated landslide-risk prediction using web GIS and machine learning models," *Sensors*, vol. 21, p. 4620, 2021.
- [3] S. L. Xu and R. Q. Niu, "Displacement prediction of Baijiabao landslide based on empirical mode decomposition and long short-term memory neural network in Three Gorges area, China," *Computers and Geosciences*, vol. 111, pp. 87–96, 2018.
- [4] S. Y. Park, H. Lim, B. Tamang et al., "A study on the slope failure monitoring of a model slope by the application of a displacement sensor," *Journal of Sensors*, vol. 2019, Article ID 7570517, 9 pages, 2019.
- [5] M. Kuradusenge, S. Kumaran, and M. Zennaro, "Rainfall-induced landslide prediction using machine learning models: the case of Ngororero District, Rwanda," *International Journal of Environmental Research and Public Health*, vol. 17, no. 11, p. 4147, 2020.
- [6] K. Yan, J. Y. Wu, Y. Q. Zhang et al., "Prediction method of deep horizontal displacement of slope soil based on damped Holt-Winters model," *Journal of Service Science and Management*, vol. 12, no. 3, pp. 391–406, 2019.
- [7] X. Z. He, H. D. Xu, H. Sabetamal, and D. Sheng, "Machine learning aided stochastic reliability analysis of spatially variable slopes," *Computers and Geotechnics*, vol. 126, p. 103711, 2020.
- [8] Y. K. Wang, H. M. Tang, T. Wen, and J. Ma, "Direct interval prediction of landslide displacements using least squares support vector machines," *Complexity*, vol. 2020, Article ID 7082594, 1155 pages, 2020.
- [9] J. W. Ma, Y. K. Wang, X. X. Niu, S. Jiang, and Z. Liu, "A comparative study of mutual information-based input variable selection strategies for the displacement prediction of seepage-driven landslides using optimized support vector regression," *Stochastic Environmental Research and Risk Assessment*, vol. 12, 2022.
- [10] Y. K. Wang, H. M. Tang, J. S. Huang, T. Wen, J. Ma, and J. Zhang, "A comparative study of different machine learning methods for reservoir landslide displacement prediction," *Engineering Geology*, vol. 298, p. 106544, 2022.
- [11] A. Ray, V. Kumar, A. Kumar, R. Rai, M. Khandelwal, and T. N. Singh, "Stability prediction of Himalayan residual soil slope using artificial neural network," *Natural Hazards*, vol. 103, no. 3, pp. 3523–3540, 2020.
- [12] K. V. Shihabudheen, G. N. Pillai, and B. Peethambaran, "Prediction of landslide displacement with controlling factors using extreme learning adaptive neuro-fuzzy inference system (ELANFIS)," *Applied Soft Computing*, vol. 61, pp. 892–904, 2017.
- [13] J. F. de Oliveira, E. G. Silva, and P. S. de Mattos Neto, "A hybrid system based on dynamic selection for time series forecasting," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 99, pp. 1–13, 2021.
- [14] W. Liu and Y. Shoji, "DeepVM: RNN-based vehicle mobility prediction to support intelligent vehicle applications," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 6, pp. 3997–4006, 2020.
- [15] M. Ma and Z. Mao, "Deep-convolution-based LSTM network for remaining useful life prediction," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 3, pp. 1658–1667, 2021.
- [16] M. Afrasiabi, M. Mohammadi, M. Rastegar, and S. Afrasiabi, "Advanced deep learning approach for probabilistic wind speed forecasting," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 1, pp. 720–727, 2021.
- [17] Y. Li, Y. Yang, K. Zhu, and J. Zhang, "Clothing sale forecasting by a composite GRU-prophet model with an attention mechanism," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 12, pp. 8335–8344, 2021.
- [18] D. Aloraifan, I. Ahmad, and E. Alrashed, "Deep learning based network traffic matrix prediction," *International Journal of Intelligent Networks*, vol. 2, pp. 46–56, 2021.
- [19] R. W. Liu, M. Liang, J. Nie, W. Y. Lim, Y. Zhang, and M. Guizani, "Deep learning-powered vessel trajectory prediction for improving smart traffic services in maritime Internet of things," *IEEE Transactions on Network Science and Engineering*, vol. 9, no. 5, pp. 3080–3094, 2022.
- [20] B. H. Ahmed and A. S. Ghabayen, "Review rating prediction framework using deep learning," *Journal of Ambient Intelligence and Humanized Computing*, vol. 13, pp. 3423–3432, 2022.

- [21] Z. J. Ma and G. Mei, "Deep learning for geological hazards analysis: data, models, applications, and opportunities," *Earth-Science Reviews*, vol. 223, p. 103858, 2021.
- [22] X. Z. Li and J. M. Kong, "Application of GA-SVM method with parameter optimization for landslide development prediction," *Natural Hazards and Earth System Sciences*, vol. 14, no. 3, pp. 525–533, 2014.
- [23] Z. Liu, W. Xu, and J. Shao, "Gauss process based approach for application on landslide displacement analysis and prediction," *Computer Modeling in Engineering & Sciences*, vol. 84, no. 2, pp. 99–122, 2012.
- [24] C. Lian, Z. Zeng, W. Yao, and H. Tang, "Ensemble of extreme learning machine for landslide displacement prediction based on time series analysis," *Neural Computing and Applications*, vol. 24, no. 1, pp. 99–107, 2014.
- [25] H. Niu, "Smart safety early warning model of landslide geological hazard based on BP neural network," *Safety Science*, vol. 123, article 104572, 2020.
- [26] H. Du, D. Song, Z. Chen, H. Shu, and Z. Guo, "Prediction model oriented for landslide displacement with step-like curve by applying ensemble empirical mode decomposition and the PSO-ELM method," *Journal of Cleaner Production*, vol. 270, p. 122248, 2020.
- [27] B. Yang, K. Yin, S. Lacasse, and Z. Liu, "Time series analysis and long short-term memory neural network to predict landslide displacement," *Landslides*, vol. 16, no. 4, pp. 677–694, 2019.
- [28] Y. Liu, C. Xu, B. Huang et al., "Landslide displacement prediction based on multi-source data fusion and sensitivity states," *Engineering Geology*, vol. 271, 2020.
- [29] 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.
- [30] Z. Liu, D. Guo, S. Lacasse, J. H. Li, B. B. Yang, and J. C. Choi, "Algorithms for intelligent prediction of landslide displacements," *Journal of Zhejiang University-Science A*, vol. 21, no. 6, pp. 412–429, 2020.
- [31] K. Zhang, K. Zhang, C. Cai, W. Liu, and J. Xie, "Displacement prediction of step-like landslides based on feature optimization and VMD-bi-LSTM: a case study of the Bazimen and Baishuihe landslides in the Three Gorges, China," *Bulletin of Engineering Geology and the Environment*, vol. 80, no. 11, pp. 8481–8502, 2021.
- [32] Z. Ma, G. Mei, E. Prezioso, Z. Zhang, and N. Xu, "A deep learning approach using graph convolutional networks for slope deformation prediction based on time-series displacement data," *Neural Computing and Applications*, vol. 33, no. 21, pp. 14441–14457, 2021.
- [33] J. Wang, G. Nie, S. Gao, S. Wu, H. Li, and X. Ren, "Landslide deformation prediction based on a GNSS time series analysis and recurrent neural network model," *Remote Sensing*, vol. 13, no. 6, p. 1055, 2021.
- [34] S. Hochreiter, "The vanishing gradient problem during learning recurrent neural nets and problem solutions," *Fuzziness and Knowledge-Based Systems*, vol. 6, no. 2, pp. 107–116, 1998.
- [35] S. Gao, Y. Huang, S. Zhang et al., "Short-term runoff prediction with GRU and LSTM networks without requiring time step optimization during sample generation," *Journal of Hydrology*, vol. 589, 2020.
- [36] H. Zhen, D. Niu, K. Wang, Y. Shi, Z. Ji, and X. Xu, "Photovoltaic power forecasting based on GA improved Bi-LSTM in microgrid without meteorological information," *Energy*, vol. 231, 2021.
- [37] Y. R. Yang, Q. U. Xiong, C. Wu et al., "A study on water quality prediction by a hybrid CNN-LSTM model with attention mechanism," *Environmental Science and Pollution Research international*, vol. 28, no. 39, pp. 55129–55139, 2021.
- [38] N. Li, L. Hu, Z. L. Deng, T. Su, and J. W. Liu, "Research on GRU neural network satellite traffic prediction based on transfer learning," *Wireless Personal Communications*, vol. 118, no. 1, pp. 815–827, 2021.
- [39] Q. Li, Z. Wang, W. Shangguan, L. Li, Y. Yao, and F. Yu, "Improved daily SMAP satellite soil moisture prediction over China using deep learning model with transfer learning," *Journal of Hydrology*, vol. 600, 2021.