Short–term displacement prediction for newly established monitoring slopes based on transfer learning

Yuan Tian, Yang–landuo Deng, Ming–zhi Zhang, Xiao Pang, Rui–ping Ma, Jian–xue Zhang

Citation: Yuan Tian, Yang–landuo Deng, Ming–zhi Zhang, Xiao Pang, Rui–ping Ma, Jian–xue Zhang, 2024. Short–term displacement prediction for newly established monitoring slopes based on transfer learning, China Geology, 7, 350–363. doi: 10.31035/cg2024053.

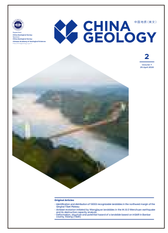
View online: https://doi.org/10.31035/cg2024053

Related articles that may interest you

China Geology 7 (2024) 350−363

China Geology

Journal homepage: http://chinageology.cgs.cn
https://www.sciencedirect.com/journal/china-geology



Short-term displacement prediction for newly established monitoring slopes based on transfer learning

Yuan Tian[a, b], Yang-landuo Deng[a, b], Ming-zhi Zhang[c, d, e,*], Xiao Pang[a, b], Rui-ping Ma[a, b], Jian-xue Zhang[a, b]

[a] *Institute of Remote Sensing and Geographical Information Systems*, *Peking University*, *Beijing* 100871, *China*
[b] *Beijing Key Laboratory of Spatial Information Integration and Its Applications*, *Beijing* 100871, *China*
[c] *Institute of Public Safety Research*, *Department of Engineering Physics*, *Tsinghua University*, *Beijing* 100084, *China*
[d] *China Institute of Geo-Environment Monitoring*, *Beijing* 100081, *China*
[e] *Technology Innovation Center for Geohazard Monitoring and Risk Early Warning*, *Ministry of Natural Resources*, *Beijing* 100081, *China*

A R T I C L E I N F O

Article history:
Received 2 April 2024
Received in revised form 19 April 2024
Accepted 22 April 2024
Available online 26 April 2024

Keywords:
Landslide
Slope displacement prediction
Transfer learning
Integrated dataset
Transformer
Pre-trained model
Universal Landslide Monitoring Program (ULMP)
Geological hazards survey engineering

A B S T R A C T

This study makes a big progress in dealing the challenges of short-term slope displacement prediction in the Universal Landslide Monitoring Program, an unprecedented disaster mitigation program in China, where lots of newly established monitoring slopes lack sufficient historical deformation data, making it difficult to extract deformation patterns and provide effective predictions which plays a crucial role in the early warning and forecasting of landslide hazards. A slope displacement prediction method based on transfer learning is therefore proposed. Initially, the method transfers the deformation patterns learned from slopes with relatively rich deformation data by a pre-trained model based on a multi-slope integrated dataset to newly established monitoring slopes with limited or even no useful data, thus enabling rapid and efficient predictions for these slopes. Subsequently, as time goes on and monitoring data accumulates, fine-tuning of the pre-trained model for individual slopes can further improve prediction accuracy, enabling continuous optimization of prediction results. A case study indicates that, after being trained on a multi-slope integrated dataset, the TCN-Transformer model can efficiently serve as a pre-trained model for displacement prediction at newly established monitoring slopes. The three-day average RMSE is significantly reduced by 34.6% compared to models trained only on individual slope data, and it also successfully predicts the majority of deformation peaks. The fine-tuned model based on accumulated data on the target newly established monitoring slope further reduced the three-day RMSE by 37.2%, showcasing great improvement in predictive accuracy. In conclusion, taking advantage of transfer learning, the proposed slope displacement prediction method effectively utilizes the available data, which enables the rapid deployment and continuous improvement of displacement predictions on newly established monitoring slopes.

1. Introduction

Landslides, defined as the movement of a mass of rock, earth or debris down a slope under the influence of gravity (Cruden D, 1991), can cause significant and irreversible loss of life, property, and ecological environment. Thus, it is vital to attach importance to monitoring, prediction and early warning of landslides.

In recent years, China has been increasing its investment in geological disaster prevention and mitigation. The Universal Landslide Monitoring Program (ULMP), an unprecedented disaster mitigation program is now carried out in mainland China (Deng L et al., 2023). In this rapidly and widely on-going program, taking the implementation of widespread monitoring instrument development and demonstration applications by the Ministry of Natural Resources in 2019 as kickoff, the National Risk Warning System on Landslides overseen by the China Institute of Geo-Environment Monitoring has deployed nearly 40000 monitoring slopes in regions such as Sichuan, Chongqing, Tibet, and Guangdong. Utilizing monitoring data for timely and effective short-term slope displacement predictions can

First author: *E-mail address*: tianyuanpku@pku.edu.cn (Yuan Tian).

* Corresponding author: *E-mail address*: zhangmingzhi@mail.cgs.gov.cn (Ming-zhi Zhang).

Literary editor: Li-qiong Jia
doi:10.31035/cg2024053

provide support for early warning and forecasting, thus having significant importance in reducing the loss and damage caused by landslide disasters.

Due to the rapidly ongoing deployment of universal landslide monitoring instruments, there are at present many newly established landslide monitoring sites with monitoring data of less than six months, nonetheless to say there will definitely be more new monitoring sites in recent future. Many of these newly established monitoring slopes have not experienced significant deformation processes at all. Their own historical data is insufficient to support reliable training based on conventional single-slope prediction models. This could lead to fatal deficiency in landslide disaster prediction and early warning, a kind of Achilles' Heel, presumably resulting in irreversible casualties and property losses. As ULMP being carried out nationwide, the number of monitoring sites will continue to increase, and the challenge of displacement prediction for newly established monitoring slopes will persist for a considerably long period. Therefore, it is absolutely necessary to design a technically feasible and effective displacement prediction method specifically for newly established monitoring slopes, enabling the rapid application and ongoing optimization of prediction models, and thus, in the long term, ensuring the achievement of the ultimate goals, efficient disaster mitigation and loss reduction, of the ULMP.

Machine learning methods possess powerful capabilities for handling non-linearity, and in recent years, have been widely applied to the prediction of slope displacement (Ma Z et al., 2021). Generally speaking, the current research trajectory for slope displacement prediction based on machine learning is as follows: (1) Decompose the cumulative slope displacement; (2) select factors influencing slope displacement; (3) establish a model for slope displacement prediction; (4) evaluate the prediction results (Tehrani FS et al., 2022). The decomposition of cumulative slope displacement into trend, cyclical, and random components proposed by Wang J (2003) and Du J et al. (2013) and others is widely used in the decomposition of slope displacement, where three components are output: The trend component represents long-term displacement, which is controlled by the internal geological conditions of the slope (such as rock type, geological structure, etc.) and reflects the macro trend of landslide development; the cyclical component is a short-term displacement influenced by various external factors (like rainfall and reservoir water levels); the random component is the displacement response caused by sudden changes in the system, which is typically ignored in actual displacement prediction work.

Existing research indicates that machine learning methods are highly effective in predicting the cyclical component of cumulative slope displacement (Tehrani FS et al., 2022). Before modeling, it is essential to conduct correlation analyses to select factors strongly correlated with cyclical displacement. Correlation analysis methods include grey relational analysis and the maximum information coefficient (Julong D et al., 1989; Reshef DN et al., 2011), among others. Once the influential factors for slope displacement are determined, suitable models are selected for modeling. Most of the modeling methods in existing research are based on supervised learning, that is, training on an existing dataset of influential factors and slope displacement, then predicting slope displacement given known influential factors. Miao F et al. (2018) used the Support Vector Regression (SVR) method to predict the cyclical displacement of slopes, while comparing with the effects of the time series and Genetic Algorithm (GA-SVR) model, Grid Search (GS-SVR), and Particle Swarm Optimization (PSO-SVR) mode. Xu S et al. (2018) selected six factors, including the previous month's cumulative rainfall and the current month's reservoir water level, to predict the cyclical displacement of slopes using the Long Short-Term Memory (LSTM) model and compared it with the SVR and Backpropagation Neural Network (BNP) models.

Existing studies on Chinese slope displacement prediction primarily focus on landslides in the Three Gorges Reservoir area or similar large-scale slopes with long-term monitoring data, conducting medium-term displacement predictions on a monthly basis (Zhang YG et al., 2021; Jiang Y et al.,2021). Research on short-term slope displacement prediction is relatively sparse. Zhu X et al. (2017) used one and a half years of historical displacement and rainfall data for daily slope displacement prediction. In their experiments, they employed the HP filter for trend-cyclical decomposition, used DES for modeling the trend component, and applied Least-Squares Support Vector Machine (LSSVM) for modeling the cyclical component. They also compared the prediction accuracy of two hyperparameter optimization methods: PSO and GA. Zhang L et al. (2020) explored the lagging effects of inducing factors on slope deformation and proposed using the Successive Projections Algorithm (SPA) method to determine the lag time of reservoir water level changes on slope deformation. Their experiments, using nearly two years of monitoring data at a 2-day interval, employed trigonometric functions for trend-cyclical decomposition and SPA-PSO-SVM for cyclical component prediction. The results showed that compared to the lag time commonly used in medium-term predictions, i.e., 1-2 months, the modeling accuracy improved after using the SPA-PSO-SVM method. Tian Y et al. (2023) combined the Temporal Convolutional Network (TCN) with the Transformer decoder, introducing a Transformer-based model, TCN-Transformer, for short-term slope displacement prediction. The TCN-Transformer was trained, verified, and tested on an eight-month-long historical landslide monitoring sequence when it took the displacement and rainfall sequences of the past 30 days as input and predicted the displacement for the next three days. Compared to traditional models like LSTM, TCN-Transformer demonstrated higher accuracy.

From the summary of existing research on slope displacement prediction, it is evident that there are significant differences in data foundation and application scenarios between previous studies and ULMP. Regarding the data foundation, most existing studies are based on long-term monitoring data for displacement prediction, with relatively abundant sets of inducing factors. In contrast, ULMP involves many newly established slopes with less than six months of monitoring data, and only rainfall data can be fully provided among all the inducing factors. In terms of application scenarios, most existing studies focus on medium-term predictions with a monthly time resolution, while ULMP requires short-term displacement prediction results, typically on a daily or even hourly basis, e. g. the current practical work demands daily deformation predictions for the upcoming three days. Therefore, existing methods in previous research are difficult to solve the problems in the context of ULMP.

Based on the background above, this study proposed a slope displacement prediction method based on transfer learning. This method could produce high-accuracy predictions of displacement on newly established monitoring slopes by leveraging generalized knowledge learned from comprehensive datasets that are relatively ample and of high quality, which, in this study, is achieved by constructing a multi-slope integrated dataset where deformation patterns from various slopes could be synthesized. This multi-slope integrated dataset could compensate for the inability of the short monitoring data sequence to express the credible slope deformation patterns. Based on this integrated dataset, pre-trained model can be established and applied to newly established monitoring slopes, enabling a rapid and reliable fulfillment of slope displacement prediction on them. Moreover, with the accumulation of slope deformation data, the model could be fine-tuned to achieve continuous optimization of the prediction performance.

2. Data and methods

2.1. Data introduction

The landslide monitoring data used in this study was sourced from the National Risk Warning System on Landslides of the China Geological Environment Monitoring Institute, the kernel system of ULMP. A total of 189 landslide monitoring slopes with 255 online instruments were selected for the research, all of which are in the rainfall-induced disaster category according to the field investigation reports in NLMP. Each of these monitoring sites is equipped with instruments like rain gauges and crack meters. And they have all been in operation for more than half a year and accumulated. relatively distinct monitoring data in one or more displacement periods, which are good data foundation for prediction model training. The chosen monitoring sites are distributed across 16 provincial administrative regions in China, with a higher concentration in Yunnan, Shaanxi, and Guizhou provinces (Fig. 1). Notably, 90% of these monitoring sites have an operational duration of less than 1 year.

2.2. Theory of transfer learning

Transfer learning is a promising machine learning concept by which prior knowledge and information from one domain and task can be applied to another domain and task (Pan SJ et al., 2009). This subsection first introduces some basic

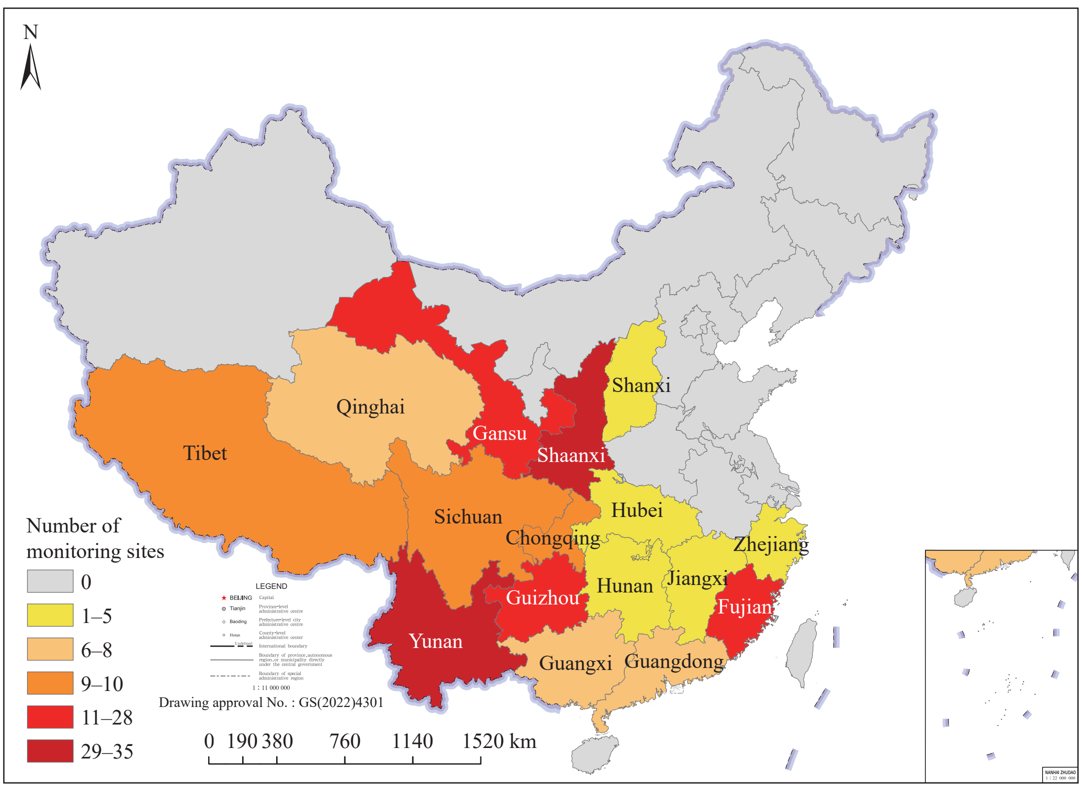


Fig. 1. Distribution map of landslide monitoring sites.

definitions and then provides a specific definition for transfer learning. The definitions are as follows (Weiss K et al., 2016):

Definition 1 (Domain):

A domain is denoted by $D=\{X,P(X)\}$, where X represents the feature space, and $P(X)$ represents the marginal distribution. In this study, the feature space usually includes historical slope displacement and rainfall data. Each parameter builds one dimension in the feature space.

Definition 2 (Task):

A task is denoted by $T=\{Y,f()\}$, where Y represents the label space, which means the space of all the label vectors. In this study, the label space only includes the slope displacement to be predicted. $f()$ represents an objective predictive function. The purpose of the objective predictive function is to predict the relationship between the features and the label, the specific task in this study is to predict the slope displacement for the next three days.

Definition 3 (Transfer Learning):

Given two different domains combined with different learning tasks, respectively, i.e., a source domain D_S and learning task T_S, and a target domain D_T and learning task T_T, the intent of transfer learning is to improve the performance of the target task in D_T by using the knowledge in D_S and T_S, where $D_S \neq D_T$ or $T_S \neq T_T$. In this study, transfer learning is adopted to improve the performance of displacement prediction in newly established monitoring slopes (D_T) by using the knowledge in predicting the displacement (T_S) at multiple slopes with relatively abundant deformation data (D_S) and how to predict their displacement.

2.3. Methodology

2.3.1. Research framework

As shown in Fig. 2, the proposed method mainly consists of four main parts:

(i) Data pre-processing, in which the original monitoring data with missing values, noise, outliers and at different intervals are interpolated, resampled and denoised.

(ii) Dataset construction, in which the data are processed into short sequences to constitute a multi-slope integrated dataset which facilitates model training and testing.

(iii) Pre-training, in which the prediction model learns deformation patterns from the source domain, i.e., multi-slope integrated dataset, which could then be applied to the prediction in target domain, i.e., newly established monitoring slopes.

(iv) Fine-tuning, in which the performance of pre-trained model is improved by freezing certain layers and training with cumulative data from target domain, i.e., newly established monitoring slopes, over time.

In short, the method mainly includes the construction of multi-slope integrated dataset, establishment and application of pre-training models, which transfer the knowledge from the source domain, i.e., multi-slope integrated dataset, to the target domain, i.e., newly established monitoring slopes, and fine-tuning based on cumulative data from the target slopes to

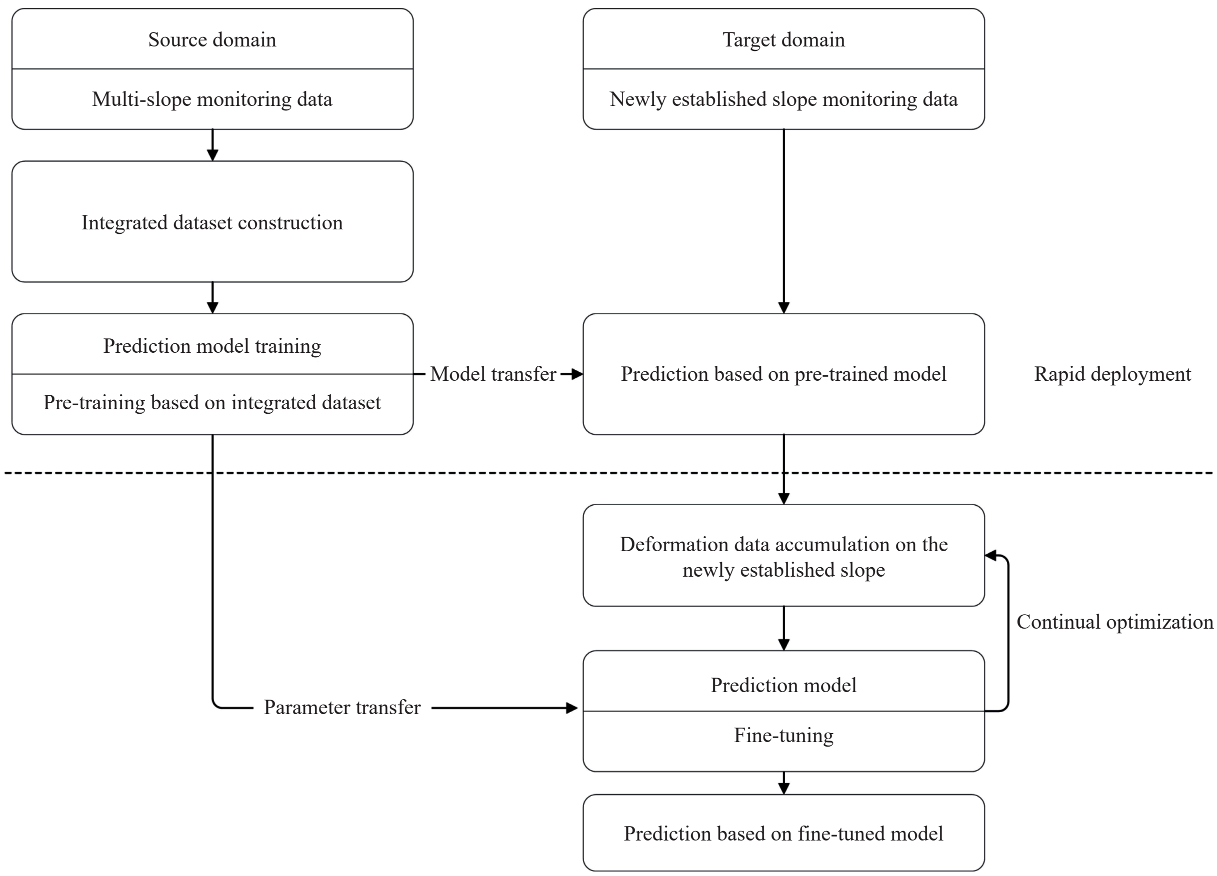


Fig. 2. Overview of transfer learning-based slope displacement prediction method.

improve the prediction performance in the target domain over time.

2.3.2. Data pre-processing

(i) Interpolation and Resampling

In the ULMP, monitoring instruments record data at regular intervals and transmit them via networks. However, data omissions frequently occur due to instrument malfunctions, network disconnections, transmission losses, or other reasons. Different types of monitoring instruments record data at different intervals, e. g. the record intervals of GNSS displacement meters and rain gauges are about 1 hour while the record intervals of ground fissure meters may vary between 5 minutes and 1 hour. In order to facilitate the following model training, the interpolation and resampling of the original monitoring data need to be done to get equispaced time series data.

Existing literature has explored how the time intervals of data influence short-term slope displacement prediction (Zhao W et al., 2020). It was discerned that a shorter resampling interval reveals more details in the curve, but it also becomes more susceptible to noise. To retain the intricacy of the curve while mitigating the impact of noise, this study chose 6-hour as the time interval for the resampling of all monitoring data, including both displacement and rainfall data. Missing values were compensated using quadratic spline interpolation.

(ii) Denoising

In this study, the Variational Mode Decomposition Partial Reconstruction (VMD-PR) method is adopted to denoise slope displacement monitoring data. Variational Mode Decomposition (VMD) is a signal decomposition method proposed by Dragomiretskiy K and Zosso D (2014). It iteratively searches for the optimal solution of variational modes, continuously updating each modal function and its central frequency, thereby decomposing the noisy signal into a series of IMF components. Vital information is mainly concentrated in the low-frequency components, while noise mainly concentrated in the high-frequency components. The Partial Reconstruction (PR) method is adopted to select low-frequency components for signal reconstruction to achieve denoising (Komaty A et al., 2012).

When VMD method is used to decompose the displacement time series, the K value is set to 8 based on the EEMD decomposition results, the penalty factor set to the default value of 2000, the noise tolerance set to 0, and the convergence criterion tolerance set to 10^{-7}. This study employs the Probability Density Function (PDF) distance to measure the similarity between the IMF and the original signal (Komaty A et al., 2013; Yang G et al., 2015), determining the number of sub-modes m used for signal reconstruction to be 5.

2.3.3. Dataset construction

In order to transfer deformation patterns of landslides to newly established monitoring slopes, a multi-slope integrated dataset with relatively abundant deformation data is to be constructed. The ULMP could provide large amount of landslide monitoring data, which help to make the construction of multi-slope integrated dataset practically feasible. The sequential data of high quality and of relatively long historical monitoring period could be provided by the National Risk Warning System on Landslides, and then integrated into the multi-slope integrated dataset.

Considering the prediction model takes an input of historical slope displacement data and rainfall data of several previous days, e. g. 60 days, and outputs the predictions of displacement for the following several days, e. g. 3 days in ULMP, it is necessary to process the monitoring data into sequences of specific length. The sliding window method is widely used to generate input samples and construct the training and testing datasets, which is illustrated in Fig. 3 (Lu H et al., 2023). The length of the original data is N, the time step of each sliding window is h , and the length of the sliding window, i.e., the length of the training sample, is Δt. The length of the sliding window should remain the same through the sampling process. As the sliding window slides forward, the original sequence of length N is divided into $M = N-(\Delta t + h) + 1$ samples (Ma J et al., 2020).

2.3.4. TCN-Transformer

The TCN-Transformer model is a short-term slope displacement prediction model based on the Transformer, formed by the combination of the Temporal Convolutional Network (TCN) and the Transformer decoder (Tian Y et al., 2023). In the context of this research, choosing the TCN-Transformer model for slope displacement prediction offers the following advantages: (1) It is designed for short-term temporal forecasting, capable of providing predictions at a daily granularity, and it exhibits high prediction accuracy, which aligns well with the requirements of ULMP efforts; (2) it is based on attention mechanisms that endows it with excellent interpretability, which may assist researchers in understanding the model's ability to capture critical information (Yanagimto H et al., 2018) and enables in-depth assessment of the model's reliability. Therefore, the TCN-Transformer model is selected as the slope displacement prediction model to be pre-trained on the multi-slope integrated dataset.

The Transformer relies entirely on the self-attention mechanism to extract intrinsic features. It can learn long-term

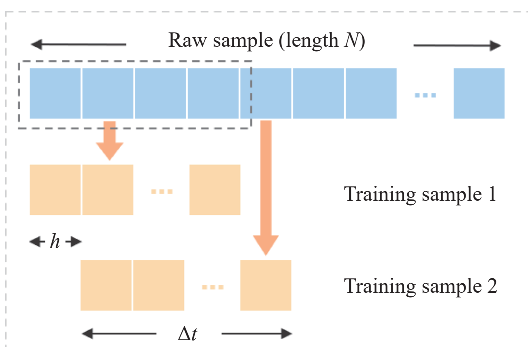


Fig. 3. Generation of sample set using the sliding window method (after Lu H et al., 2023).

dependencies and is more easily parallelizable compared to RNN-based models. The original Transformer, applied to classification tasks, acquires high-dimensional features for single-timepoint samples via an embedding layer and uses positional encoding to retain the position information in the original sequence. Subsequently, based on the query-key-value concept, the self-attention layer maps the original features to Q, K, and V feature sets. It uses Q and K to calculate the relevance of features at different timepoints with the current timepoint and obtains the features for the next timepoint by a weighted sum of V (Vaswani A et al., 2017), with a formula for calculating the weight of the t−k feature at timepoint t as follows (Equ. 1):

$$weight_{t-k} = softmax\left(Q_t \times K^T\right)_{t-k}, \quad k = 0, 1, 2, \ldots, t \tag{1}$$

However, slope displacement prediction is in essence a regression problem, and the traditional embedding layer in Transformer is not suitable for feature extraction in regression tasks. Considering that a single timepoint sample cannot effectively describe the current state of the landslide's motion, Tian Y et al. (2023) adopts the Temporal Convolutional Network (TCN) to replace the embedding layer. By using TCN, it is possible to introduce controllable length of historical information into the current timepoint's features without changing the input format, while retaining the parallelizability of the Transformer model training.

The overall structure of the TCN-Transformer model used in this study is illustrated in Fig. 4 (Tian Y et al., 2023). Sequence data including slope displacement and rainfall are processed through the TCN. Features at corresponding timepoints are then concatenated. After processed through a fully connected layer for feature mixing and appended with positional encoding, they are input into the self-attention layer to output the current timepoint's hidden state h_t. Finally, the slope displacement of the next timepoint is produced through a fully connected layer.

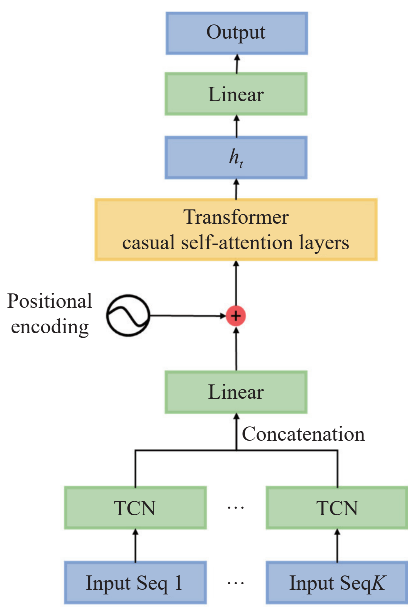


Fig. 4. TCN-Transformer for short-term slope displacement predictions (after Tian Y et al., 2023).

2.3.5. Pre-training

When landslide monitoring sites have just been established, there is a paucity of historical monitoring data, especially data from deformation periods. As a result, it becomes challenging to learn landslide deformation patterns from such limited data. Blindly applying single-slope modeling methods in these scenarios could potentially lead to misreporting or underreporting of landslide hazards, resulting in irreversible losses to life and property.

In response to this challenge, transfer learning techniques are employed to fully leverage the model's learning capacity, which involves transferring knowledge or patterns learned from a well-populated source domain to the target domain. Essentially, this means migrating the parameters of the prediction model pre-trained on the multi-slope integrated dataset to the prediction model of the newly established monitoring slope. This procedure equates to using the pre-trained model, trained on the multi-slope integrated dataset, to predict the displacement of the new slope. In this study, the TCN-Transformer is employed as the pre-trained model, the results of which will be evaluated.

2.3.6. Fine-tuning

Using a pre-trained model enables rapid fulfillment of prediction on newly established monitoring slopes. When substantial deformation data being acquired sooner or later in the ongoing monitoring process, it becomes feasible to fine-tune the pre-trained model using the actual accumulated data. This process aims to further improve the prediction performance continually on the newly established monitoring slopes. As the monitoring data from the newly established monitoring slopes continue to grow, fine-tuning iterations may successively optimize the model. This process provides a practical and effective method to improve slope displacement prediction outcome for the newly established monitoring slopes in ULMP.

According to previous research on fine-tuning, the initial layer in the deep neural network focus on capturing more general and basic features while the deeper layers are usually aimed to extract more special and complex features (Howard J et al., 2018). As shown in Fig. 5 (Lemley J et al., 2017), when transferring parameters from the pre-trained model A, the parameters of the first layer could be retained while the remaining layers could be fine-tuned to meet specific task scenarios. This process could be conducted by freezing the parameters of the initial layer of the deep learning model. In this study, the initial TCN layer of the TCN-Transformer model is frozen and the output layer to be fine-tuned. Whether the attention layer is to be fine-tuned or not should depend on the experiment results.

3. Results and analysis

3.1. Dataset construction

After the above-mentioned data pre-processing step, 255

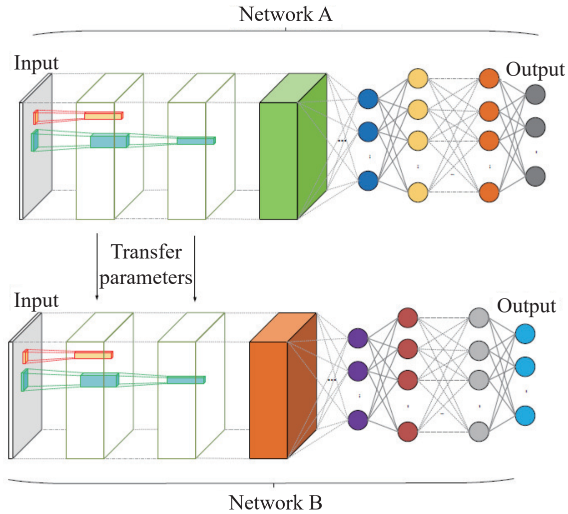


Fig. 5. Pre-trained model parameter transfer method (after Lemley J et al., 2017).

high-quality monitoring sequence data with a time resolution of 6 hours are obtained. The prediction model takes an input of historical slope displacement data and rainfall data of 60 days and outputs the predictions of displacement for the next 3 days. Therefore, when constructing the dataset, it is necessary to process the monitoring data into sequences of 60 days. The sliding window method is employed on the original monitoring sequences, setting the time step of each sliding window h to 3, i.e., 18 h, and the length of sliding window Δt to 240, i.e., 60 days. This results in an integrated dataset of 23014 sequences, each with a length of 240. Each sequence comprises two dimensions, displacement values and rainfall data. From this dataset, 2067 sequences are randomly selected, originating from 11 monitoring instruments, to serve as the test set. The remaining 20947 sequences were designated as the training set, i.e., the multi-slope integrated dataset.

3.2. Evaluation indicators

In this study, the Root Mean Square Error ($RMSE$) and coefficient of determination (R^2) were used to evaluate the predictive performance of the model (Huang F et al., 2017; Wang Y et al., 2022). The formulas for $RMSE$ and R^2 are given by Equ. (2) and Equ. (3) respectively, where y_t and $\hat{y}_t$ represent the true values and predicted values at time t, $\bar{y}$ is the mean of the true value sequence, and n is the length of the sequence.

$$RMSE = \sqrt{\frac{\sum_{t=1}^{n}\left(\hat{y}_t - y_t\right)^2}{n}} \tag{2}$$

$$R^2 = \frac{\sum\left(\hat{y}-\bar{y}\right)^2}{\sum\left(y-\bar{y}\right)^2} = 1 - \frac{\sum\left(y-\hat{y}\right)^2}{\sum\left(y-\bar{y}\right)^2} \tag{3}$$

$RMSE$ can represent the average magnitude of errors between predicted values and true values. The smaller the $RMSE$ value, the smaller the prediction error and the higher the prediction accuracy of the model. $RMSE$ is greatly influenced by the numerical scale of the predicted values and the true values, and cannot fully capture the morphological relationship between the prediction value curve and the true value curve. However, in practical work, the prediction accuracy of the rapid deformation stage of landslides is often more important. Therefore, R^2 is also introduced as an evaluation index in this study. R^2 can represent the degree of fitting between predicted values and true values. The more similar the shape of the predicted value curve is to the true value curve, the closer R^2 is to 1, which can supplement the deficiency of $RMSE$ index.

3.3. Model pre-training and result evaluation

As depicted in the Methodology section, the TCN-Transformer model is employed as the pre-trained model. The input sequence length for the model was set to 60 days. Given the time resolution of the sequence is 6 hours, the displacement sequence was split into samples of length 240. Specific model parameters are as follows:

(1) For the TCN part, the number of convolutional layers is set to 2 with a kernel size of 9 and an output channel count of 32. The final features consider information from the previous 16 time points, representing the past 4 days of sample data. (2) The attention mechanism has 1 layer with a feature size of 64 and utilizes 4 attention heads.

During model training, the L1Loss was used as the loss function. The initial learning rate was set to 10^{-4}, and the Adam optimization technique was employed for training (Kingma DP et al., 2014). The ratio between the training set and the validation set was maintained at 9∶1.

To evaluate the pre-training model's prediction performance on newly established landslide monitoring slopes, the pre-trained model was trained on the training set, and then tested on the test set. In order to assess the pre-training model's overall performance, it is also necessary to employ the TCN-Transformer model for single-slope model training on the test set, i.e., 11 monitoring instruments, separately and to compare the performance of these 11 single-slope models with those of the pre-training models.

During single-slope modeling, historical monitoring data from each slope must be used for training. In the experiments, the monitoring data of the first 90 days, or almost 3 months, were extracted as training samples to simulate the situation where historical data is insufficient for newly established monitoring slopes. The remaining data of each slope in the test set was used as the real values to observe the models' ability to predict future slope deformations, thus allowing for a performance comparison between the predictions made by the pre-trained model and single-slope model on the test set. When training the single-slope models, the ratio between the training set and the validation set was also set at 9∶1.

Table 1 presents the average prediction accuracy of the single-slope model and the pre-trained model on the test set, from the beginning of the 91st day as stated above. The three-

Table 1. Average accuracy of displacement prediction on test set.

	DAY1		DAY2		DAY3	
	RMSE/mm	R^2	*RMSE*/mm	R^2	*RMSE*/mm	R^2
Single slope model	13.577	−0.05	13.735	−0.1	14.012	−0.203
Pre-trained model	7.71	0.652	9.064	0.536	10.268	0.445

day prediction results of the pre-trained model on the test set have gained $RMSE$ values of 7.710 mm, 9.064 mm, and 10.268 mm, respectively, and R^2 values of 0.652, 0.536, and 0.445, respectively. In all metrics, the pre-trained model outperforms the single-slope model. The average three-day $RMSE$ for the pre-trained model is 9.014 mm, a 34.6% reduction from the single-slope model's average $RMSE$ of 13.775 mm. The R^2 values of the single-slope model's three-day predictions are all less than 0, indicating a prediction performance worse than the average and implying that its predictions are essentially unreliable, while all the R^2 values of the pre-trained model's three-day predictions are positive.

The primary reason for this discrepancy could be the unevenly distribution of different deformation motion states in slopes, e.g., many slopes remain almost still in most time. Strictly, in supervised learning, models can only learn through the information present in the training set. If the displacement data in the training set cannot adequately represent deformation patterns, then the trained model is also incapable of making accurate extrapolations. Based on a limited number of training samples further compromised by an imbalance in slope displacement motion states, i.e., samples in stable state vastly outnumbering those in a deforming state, the models might have resulted in a severe underrepresentation of the slope's deforming states. This inadequacy would more likely lead to the failure of the single-slope modeling approach where the data are much smaller than the integrated dataset.

Considering that the $RMSE$ metric has a strong correlation with the displacement scale of the slope itself, it may not fully reflect the differences in prediction performance across various slopes. Therefore, this study adopts the R^2 distribution pattern to further analyze the predict performance of the models.

Fig. 6 presents the box plot of the three-day prediction results in terms of R^2. It is evident that the R^2 values of the single-slope models are generally distributed around the 0 value, indicating poor prediction performance. In contrast, the pre-trained model exhibits a more concentrated R^2 distribution and the upper quartiles for the three-day predictions are 0.866, 0.711, and 0.613, respectively, showing that the pre-trained models are significantly superior to the single-slope models on the test set. It is commonly agreed that accurate predictions of peak deformation values are practically important for efficient landslide warning (Tian Y et al., 2023). To further observe the practical prediction performance of the pre-trained models, a slope, Yinjiaao slope, is randomly singled out from the test set to illustrate the results in more details. As depicted in Fig. 7, the prediction values for the first day shows an overestimation for the first peak, which could be reasonably more welcomed that a shortage in prediction value, and although the magnitude of the prediction for the second peak is relatively accurate, there is a slight temporal lag. In case of the second- and third-day predictions, the predicted values for the first peak are obviously smaller than the true values although the predicted values for the second peak are much better.

The three-day prediction results for the Yingshang Kaziping slope, which has three deformation peaks, are shown in Fig. 8. It provides a relatively accurate prediction of the timing for two big deformation processes when the predicted values of the peak are generally high. Specifically, the forecast for the first day convincingly predicts a minor deformation event occurring around August 5th. This suggests that the model's predictions have practical value in the early warning and forecasting of landslides.

Overall, even though there are still discrepancies between the predicted and actual deformation peak values using the pre-trained models, pre-trained models are proven capable of producing moderately accurate predicting of the timings of

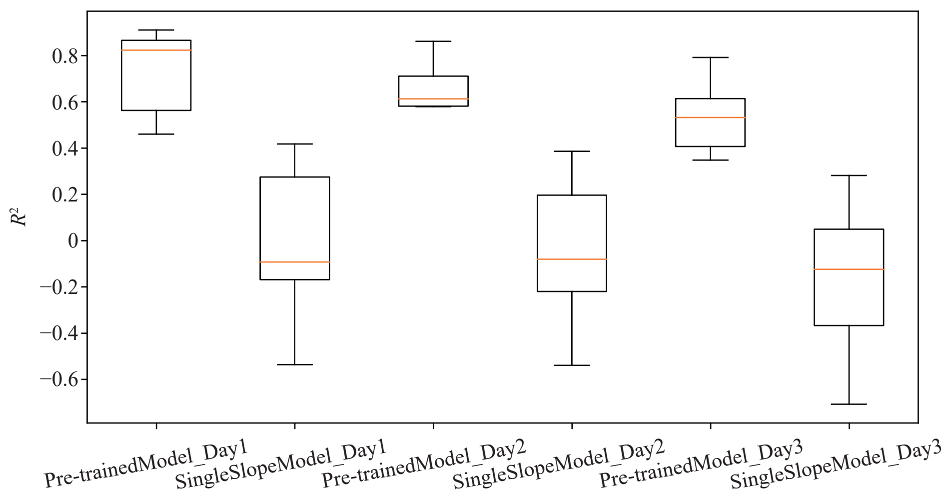


Fig. 6. R^2 box diagram for 3-day displacement prediction on the test set.

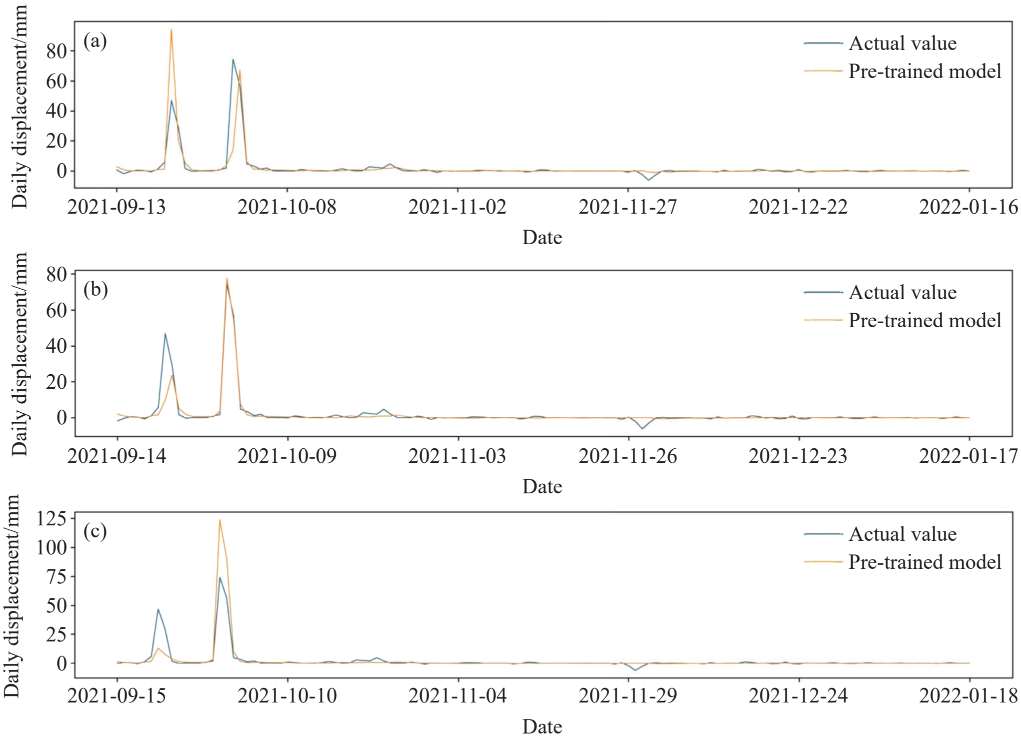


Fig. 7. Three-day prediction results of Yinjiaao slope by pre-trained model. a–Day 1; b–Day 2; c–Day 3.

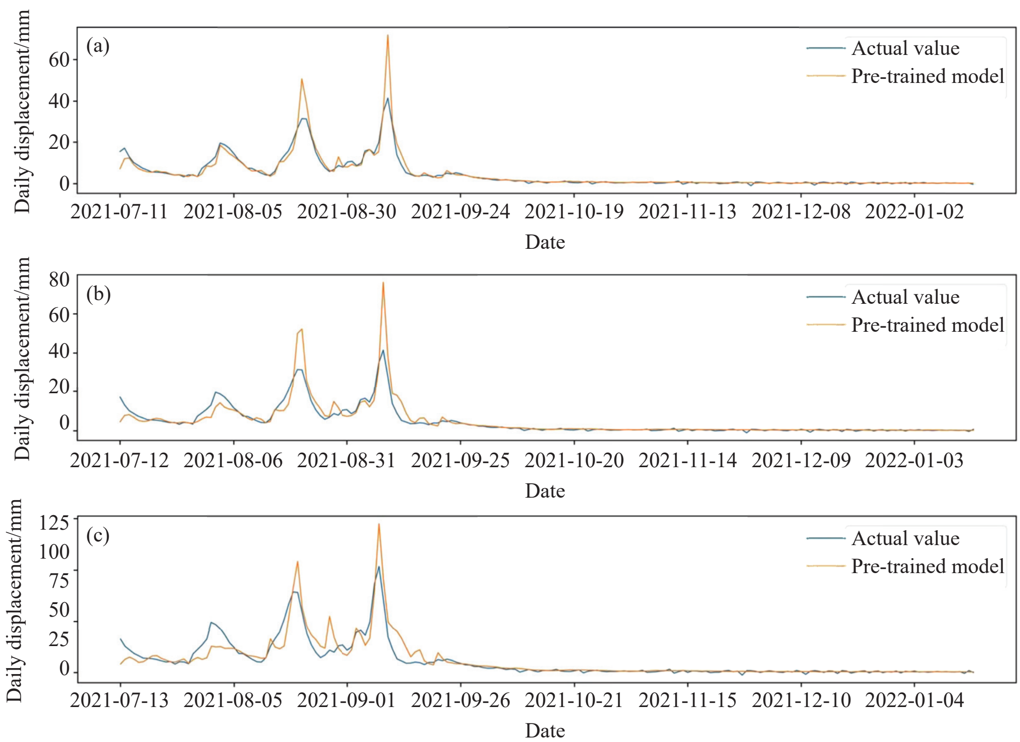


Fig. 8. Three-day prediction results of Yingshang Kaziping slope by pre-trained model. a–Day 1; b–Day 2; c–Day 3.

deformation peaks on newly established monitoring slopes, which can offer valuable support for early warning and forecasting of landslide disasters in practice.

3.4. Model fine-tuning and result evaluation

In Section 3.3, it is observed that the Yingshang Kaziping slope underwent multiple deformations in a single flood season. In this section, monitoring data from this slope are used for single-slope fine-tuning experiments to compare the prediction effects of the fine-tuned model with the prediction effects of the pre-trained model. Fig. 9 shows the displacement and rainfall monitoring data for the Yingshang Kaziping slope. It is obvious that significant rainfalls preceded the four deformation peaks. After the rain subsided, the slope deformation gradually receded to stability. There is a clear correlation between the slope deformations and rainfall, with the overall deformation pattern resembling that of stepwise landslides.

For the fine-tuning experiment, the starting point of the last rapid deformation, August 25, 2021, is taken as a split point. Monitoring data from April 15 to August 25, 2021, are used for model training, with which a deformation dataset from August 8 to August 25, 2021, are extracted as a validation set, while the remainder serving as the training set. Data after August 25, 2021, are extracted as the test set.

The model framework and parameter settings for the single-slope fine-tuning are consistent with the pre-trained model constructed in Section 3.3. Besides fine-tuning the output layer, i.e., fully connected layer, the experiment also evaluated the results of fine-tuning of the self-attention component. During model training, the L1Loss is employed as the loss function, with an initial learning rate set to 10^{-5}. The model is trained using the Adam optimization method. Fig. 10 depicts the loss curves on the validation set. It is evident that the convergence is fairly good and the loss value obviously reduced when both the self-attention component and output layer are fine-tuned. In contrast, when only the output layer is fine-tuned, due to fewer parameters adjustable, the change in loss value was minimal, resulting in subpar training results. Thus, for this experiment, both the self-attention and fully connected layers are to be fine-tuned.

Table 2 presents the prediction accuracy of the pre-trained and fine-tuned models on the test set. After fine-tuning the pre-trained model, the single-slope fine-tuned model's three-day prediction results gain $RMSE$ values of 1.248 mm, 1.773 mm, and 1.976 mm, R^2 values of 0.955, 0.910, and 0.887, respectively, obviously superior to the pre-trained model. The average three-day prediction $RMSE$ of the single-slope fine-tuned model is 1.666 mm, a 37.2% decrease from the pre-trained model's average $RMSE$ of 2.653 mm, demonstrating a significant improvement in prediction accuracy through fine-tuning.

Fig. 11 displays the prediction results of both the pre-trained and fine-tuned models on the test set. While the overall trends of the three-day predictions from both models are very similar, it is still obvious that the single-slope fine-tuned model's predictions are closer to the actual values compared to the pre-trained model. Taking a closer inspection at the rapid deformation period, from August 25 to September 24, the fine-tuned model's predictions align better with the actual curve than output of the pre-trained model, with the

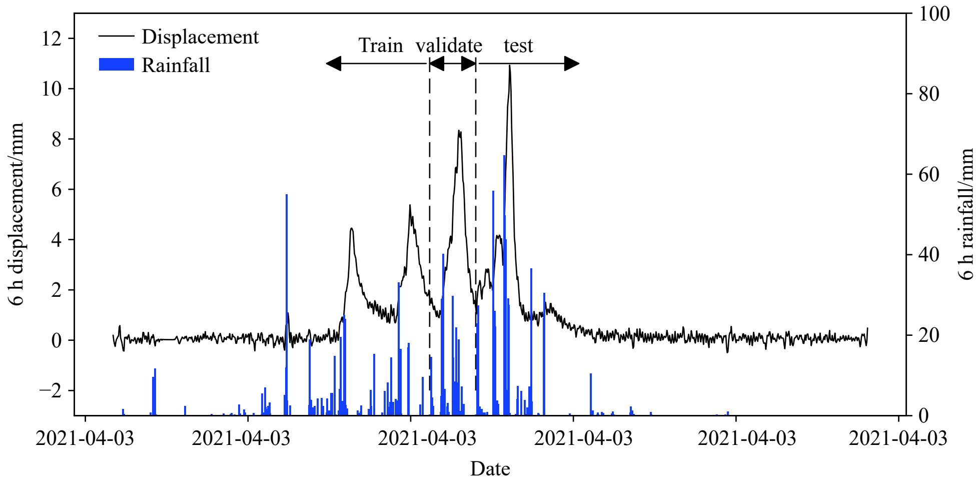


Fig. 9. Experiment data for fine-tuning.

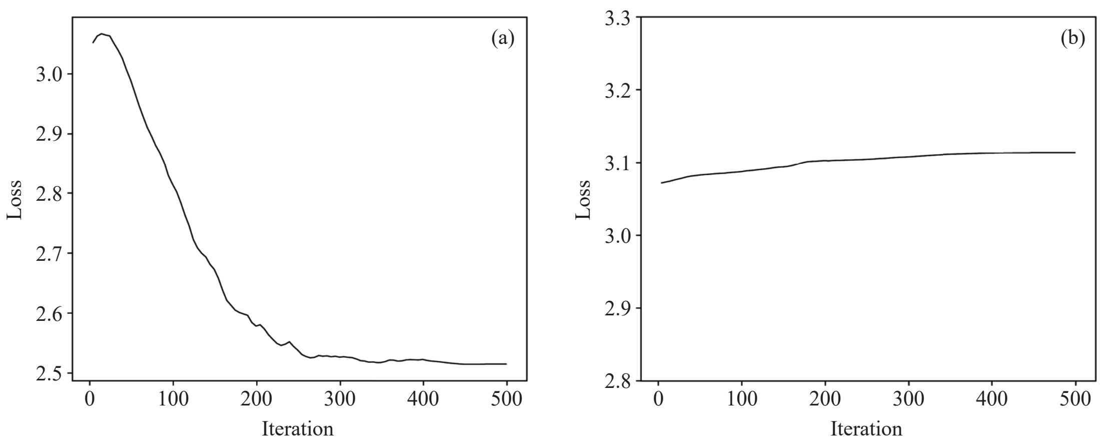


Fig. 10. Single slope model fine-tuning loss curve. a–Fine-tune the attention and output layers; b–Fine-tune the output layer only.

predicted deformation peaks being much closer to the real values. It is noticeable that after single-slope fine-tuning, the issue of over-estimating peak values in the pre-trained model, also seen in Section 3.3, is markedly mitigated.

4. Discussion

4.1. Assessment of the reliability of the method

The TCN-Transformer model possesses the capability to output self-attention. Within the self-attention mechanism, the attention weights at the current time step, derived from factors across different time steps, reflect the model's attention distribution towards historical information. This can aid researchers in assessing the model's credibility (Yanagimto H et al., 2018). In this study, modifications are made to the fast-transformer software package, incorporating Q and K into the model's output. Based on Equ. (1), the model's attention distribution can be computed.

The attention distribution of the single-slope fine-tuned model during rapid deformation periods is illustrated in Fig. 12. It can be observed that the attention in head 1 is predominantly centered around the displacement peaks near August 20 and September 8, while the distribution in head 2, 3, and 4 appears relatively uniform, without any obvious dominant focus.

Fig. 13 depicts the temporal attention distribution of all the four attention heads. Overall, the model's attention is primarily concentrated within the preceding 20 days. Specifically, the attention in head 2, 3, and 4 is mainly distributed within the initial 4 days, while the attention in head 1 exhibits a two-bell-shaped distribution mainly focusing on 0-5 days and 8-17 days, corresponding with the peak-centered distribution characteristics seen in Fig. 12. Collating the above data, it is evident that head 1 mainly focuses on critical events like rapid deformations, possessing the ability to extract long-distance information, whereas head 2, 3 and 4 primarily attend to the data information of the recent 4 days.

In summary, according to the resulting attention distribution, the single-slope fine-tuning model can effectively integrate long-term and short-term information from historical monitoring data. It pays special attention to key events like displacement peaks and extrapolates the patterns of slope displacement fairly accurately and, therefore, appears to have great rationality and credibility.

Table 2. Average accuracy of displacement prediction at Yingshang Kaziping slope.

	DAY1		DAY2		DAY3	
	RMSE/mm	R^2	*RMSE*/mm	R^2	*RMSE*/mm	R^2
Pre-trained model	2.193	0.862	3.099	0.724	2.666	0.795
Fine-tuned model	1.248	0.955	1.773	0.91	1.976	0.887

4.2. Limitations and outlook

According to the engineering practice in ULMP, the frequency of Fine-tuning optimization could be set to once a month. However, further experiments could be conducted to determine the optimal frequency, taking into account both the optimization effect and available computing resources comprehensively.

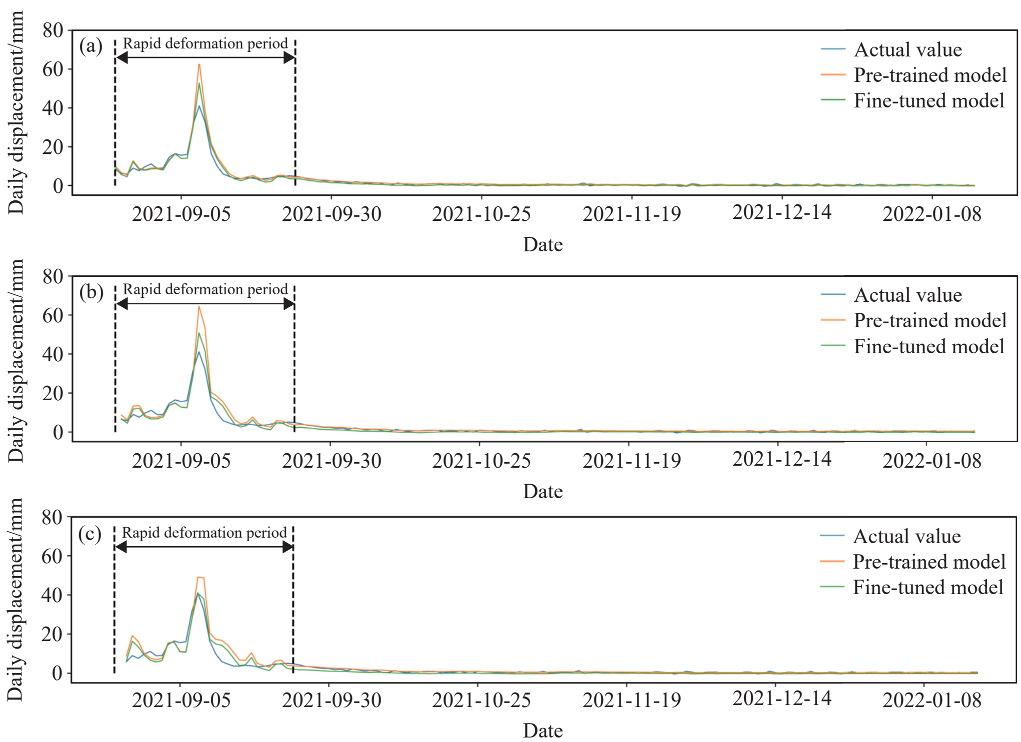


Fig. 11. Displacement prediction results of Yingshang Kaziping slope. a–Day 1; b–Day 2; c–Day 3.

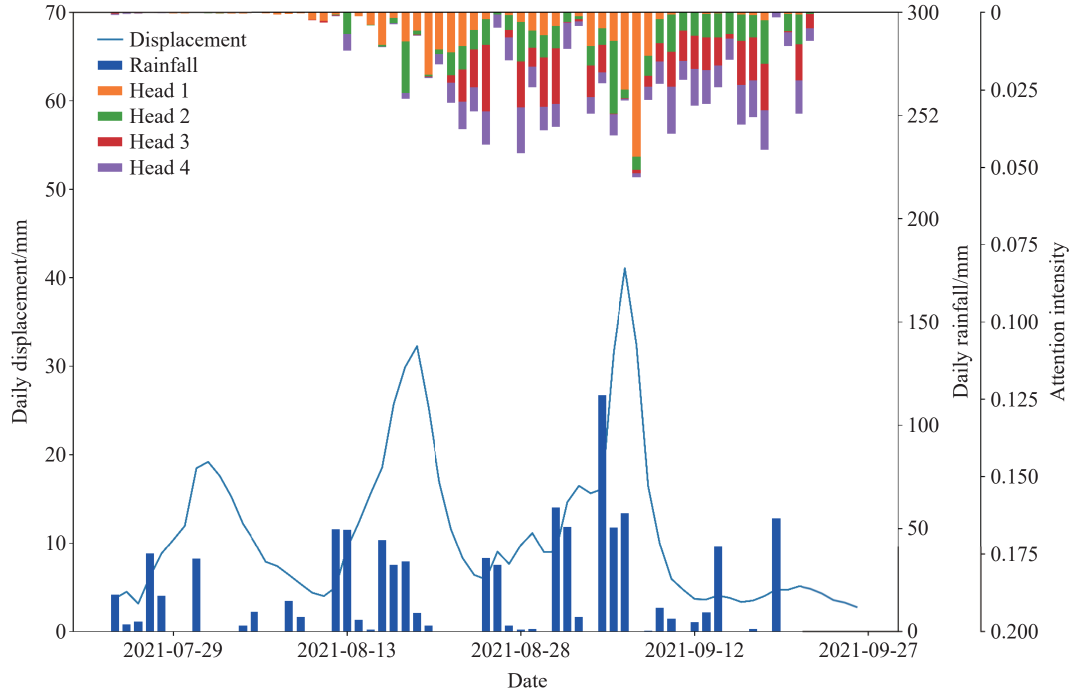


Fig. 12. Attention distribution of single-slope fine-tuning model.

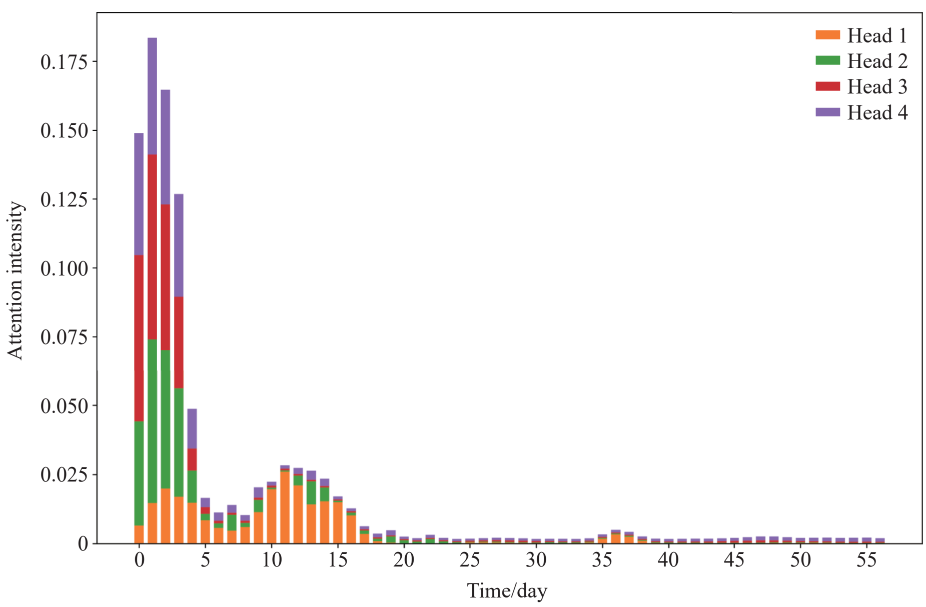


Fig. 13. Distribution of attention with time distance of single-slope fine-tuning model.

Considering that slopes of different geological types often exhibit varying deformation patterns, it is reasonable and potential to consider classifying slopes based on their geological type and conducting transfer learning in different groups. By acquiring the geological information of the newly established monitoring slopes in the future, further enhancement in prediction performance may be achieved.

5. Conclusion

This study proposes a novel landslide displacement prediction method based on transfer learning. The method utilizes a multi-slope integrated dataset for model pre-training and supports subsequent fine-tuning optimization, enabling rapid deployment and continuous optimization of deformation prediction for newly established monitoring slopes.

(i) Pre-training the model on the constructed multi-slope integrated dataset effectively compensates for the deficiency of short-term data from newly established slopes in expressing deformation patterns. It facilitates the rapid deployment and application of displacement prediction models at new monitoring sites. Case studies demonstrate that the TCN-Transformer model, trained on the multi-slope integrated dataset, can be used as a pre-trained model for displacement prediction at new monitoring slopes. The three-day average RMSE decreases by 34.6% compared to single-slope models, and the model successfully predicts most deformation peaks.

(ii) After accumulating a certain amount of deformation monitoring data at new monitoring sites, transferring the pre-trained model trained on the comprehensive dataset from multiple slopes to a single slope for fine-tuning effectively improves the model's prediction accuracy. With continuous accumulation of deformation data, the displacement prediction model can be continuously optimized. In the case studies, the three-day prediction *RMSE* of the fine-tuned single-slope model is 1.248 mm, 1.773 mm, and 1.976 mm, with R^2 values of 0.955, 0.910, and 0.887, respectively, indicating a significant improvement in prediction accuracy compared to the pre-trained model.

(iii) According to the further inspection of the attention distribution output from the model of Yingshang Kaziping slope, it was observed that the attention is also concentrated around key events such as peak displacements. This indicates the model's high rationality and credibility.

In summary, the proposed method effectively addresses the key issue of existing landslide displacement prediction models being challenging to apply to newly established monitoring slopes in ULMP. It could help reducing the loss caused by landslide disasters and protect the security of people's lives and property.

CRediT authorship contribution statement

Yuan Tian: Design of this modelling study and draft of the manuscript. Yang-landuo Deng: Modelling experiments, results analysis, and interpretation. Ming-zhi Zhang: Scientific question choice, data collection, paper revising, and discussion. Xiao Pang and Rui-ping Ma: Codes writing, debugging, and modification. Jian-xue Zhang: Data preprocess and discussion.

Declaration of competing interest

The authors declare no conflict of interest

Acknowledgment

This work was funded by the project of the China Geological Survey (DD20211364), the Science and Technology Talent Program of Ministry of Natural Resources of China (grant number 12110600000018003 9–2201).

References

Cruden D. 1991. A simple definition of a landslide. Bulletin of Engineering Geology & the Environment, 43(1), 27–29. doi: 10.1007/BF02590167.

Deng L, Yuan H, Zhang M, Chen J. 2023. Research progress on landslide deformation monitoring and early warning technology. Journal of Tsinghua University (Science and Technology), 63(6), 849–864 (in Chinese with English abstract). doi: 10.16511/j.cnki.qhdxxb.2023.22.002.

Dragomiretskiy K, Zosso D. 2014. Variational mode decomposition. IEEE transactions on signal processing, 62(3), 531–544. doi: 10.1109/TSP.2013.2288675.

Du J, Yin K, Lacasse S. 2013. Displacement prediction in colluvial landslides, three Gorges reservoir, China. Landslides, 10, 203–218. doi: 10.1007/s10346-012-0326-8.

Howard J, Ruder S. 2018. Universal language model fine-tuning for text classification. arXiv preprint arXiv: 1801.06146. doi: 10.48550/arXiv.1801.06146.

Huang F, Huang J, Jiang S, Zhou C. 2017. Landslide displacement prediction based on multivariate chaotic model and extreme learning machine. Engineering Geology, 218, 173–186. doi: 10.1016/j.enggeo.2017.01.016.

Jiang Y, Xu Q, Lu Z, Luo H, Liao L, Dong X. 2021. Modelling and predicting landslide displacements and uncertainties by multiple machine-learning algorithms: Application to Baishuihe landslide in Three Gorges Reservoir, China. Geomatics, Natural Hazards and Risk, 12(1), 741–762. doi: 10.1080/19475705.2021.1891145.

Julong D. 1989. Introduction to grey system theory. The Journal of Grey System, 1(1), 1–24.

Kingma DP, Ba J. 2014. Adam: A method for stochastic optimization. arXiv preprint arXiv: 1412.6980. doi: 10.48550/arXiv.1412.6980.

Komaty A, Boudraa A, Dare D. 2012. EMD-based filtering using the Hausdorff distance. 2012 IEEE International Symposium on Signal Processing and Information Technology, ISSPIT, 292–297. doi: 10.1109/ISSPIT.2012.6621303.

Komaty A, Boudraa AO, Augier B, Daré-Emzivat, D. 2013. EMD-based filtering using similarity measure between probability density functions of IMFs. IEEE Transactions on Instrumentation and Measurement, 63(1), 27–34. doi: 10.1109/TIM.2013.2275243.

Lemley J, Bazrafkan S, Corcoran P. 2017. Transfer learning of temporal information for driver action classification. MAICS, 123–128.

Lu H, Wu J, Ruan Y, Qian F, Meng H, Gao Y, Xu T. 2023. A multi-source transfer learning model based on LSTM and domain adaptation for building energy prediction. International Journal of Electrical Power & Energy Systems, 149, 109024. doi: 10.1016/j.ijepes.2023.109024.

Ma J, Li Z, Cheng JC, Ding Y, Lin C, Xu Z. 2020. Air quality prediction at new stations using spatially transferred bi-directional long short-term memory network. Science of The Total Environment, 705, 135771. doi: 10.1016/j.scitotenv.2019.135771.

Ma Z, Mei G, Piccialli F. 2021. Machine learning for landslides prevention: A survey. Neural Computing and Applications, 33, 10881–10907. doi: 10.1007/s00521-020-05529-8.

Miao F, Wu Y, Xie Y, Li Y. 2018. Prediction of landslide displacement with step-like behavior based on multialgorithm optimization and a support vector regression model. Landslides, 15, 475–488. doi: 10.1007/s10346-017-0883-y.

Pan SJ, Yang Q. 2009. A survey on transfer learning. IEEE Transactions on knowledge and data engineering, 22(10), 1345–1359. doi: 10.1109/TKDE.2009.191.

Reshef DN, Reshef YA, Finucane HK, Grossman SR, McVean G, Turnbaugh PJ, Lander ES, Mitzenmacher M, Sabeti PC. 2011. Detecting novel associations in large data sets. Science, 334(6062), 1518–1524. doi: 10.1126/science.1205438.

Tehrani FS, Calvello M, Liu Z, Zhang L, Lacasse S. 2022. Machine learning and landslide studies: Recent advances and applications. Natural Hazards, 114(2), 1197–1245. doi: 10.1007/s11069-022-05423-7.

Tian Y, Pang X, Zhao W, Chang X, Cheng C, Zou P, Cao X. 2023. A transformer-based model for short-term landslide displacement prediction. Acta Scientiarum Naturalium Universitatis Pekinensis, 59(2), 197–210. 10.13209/j. 0479-8023.2022. 106. (in Chinese)

Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser L, Polosukhin, I. 2017. Attention is all you need. In:

Proceedings of the 31st International Conference on Neural Information Processing Systems, 6000–6010

Wang J. 2003. Quantitative prediction of landslide using S-curve. The Chinese Journal of Geological Hazard And Control, 14, 1–8.

Wang Y, Tang H, Huang J, Wen T, Ma J, Zhang J. 2022. A comparative study of different machine learning methods for reservoir landslide displacement prediction. Engineering Geology, 298, 106544. doi: 10.1016/j.enggeo.2022.106544.

Weiss K, Khoshgoftaar TM, Wang DD. 2016. A survey of transfer learning. Journal of Big data, 3(1), 1–40. doi: 10.1186/s40537-016-0043-6.

Xu S, Niu R. 2018. Displacement prediction of Baijiabao landslide based on empirical mode decomposition and long short-term memory neural network in Three Gorges area, China. Computers & Geosciences, 111, 87–96. doi: 10.1016/j.cageo.2017.10.013.

Yanagimto H, Hashimoto K, Okada M. 2018. Attention visualization of gated convolutional neural networks with self attention in sentiment analysis. 2018 International Conference on Machine Learning and Data Engineering (iCMLDE), IEEE, 77–82. doi: 10.1109/iCMLDE.2018.00024.

Yang G, Liu Y, Wang Y, Zhu Z. 2015. EMD interval thresholding denoising based on similarity measure to select relevant modes. Signal Processing, 109, 95–109. doi: 10.1016/j.sigpro.2014.10.038.

Zhang L, Shi B, Zhu H, Yu XB, Han H, Fan X. 2020. PSO-SVM-based deep displacement prediction of Majiagou landslide considering the deformation hysteresis effect. Landslides, 18, 179–193. doi: 10.1007/s10346-020-01426-2.

Zhang YG, Tang J, He ZY, Tan J, Li C. 2021. A novel displacement prediction method using gated recurrent unit model with time series analysis in the Erdaohe landslide. Natural Hazards, 105, 783–813. doi: 10.1007/s11069-020-04337-6.

Zhao W, Ma J, Zhang M, Cheng C, Pang X, Qi G. 2020. Effects of sampling frequency on short-term prediction of landslide displacement: A case study of Kamenziwan landslide. In: Proceedings of the 2020 2nd International Conference on Big Data and Artificial Intelligence, 144–148. doi: 10.1145/3436286.3436471.

Zhu X, Xu Q, Tang M, Nie W, Ma S, Xu Z. 2017. Comparison of two optimized machine learning models for predicting displacement of rainfall-induced landslide: A case study in Sichuan Province, China. Engineering geology, 218, 213–222. doi: 10.1016/j.enggeo.2017.01.022.

Zhuang F, Qi Z, Duan K, Xi D, Zhu Y, Zhu H, Xiong H, He Q. 2020. A comprehensive survey on transfer learning. Proceedings of the IEEE, 109(1), 43–76. doi: 10.1109/JPROC.2020.3004555.