

Article

Displacement Prediction and Monitoring Methods for Baishui River Landslide in the Three Gorges Reservoir Area

Jiayan Yin ^{1,†}, Jiachuang Song ^{2,†}, Kai Xie ^{1,*}, Hongling Tian ³ , Jianbiao He ⁴ and Wei Zhang ⁵

¹ School of Electronic Information and Electrical Engineering, Yangtze University, Jingzhou 434023, China; 2024720741@yangtzeu.edu.cn

² School of Future Technology, Yangtze University, Jingzhou 434023, China; 2024001622@yangtzeu.edu.cn

³ Institute of Mountain Hazards and Environment, Chinese Academy of Sciences, Chengdu 610041, China; thl@imde.ac.cn

⁴ School of Computer Science, Central South University, Changsha 410083, China; jbh@mail.csu.edu.cn

⁵ School of Electronic Information, Central South University, Changsha 410004, China; csuzwzbn@csu.edu.cn

* Correspondence: xiekai@yangtzeu.edu.cn

† These authors contributed equally to this work.

Abstract

Predicting landslide displacement is important for geological-hazard early warning. In reservoir areas, displacement evolution is affected by rainfall, reservoir water level, vegetation variation, and the intrinsic non-stationarity of the displacement sequence, which makes accurate prediction difficult for conventional single-sequence models. To address this problem, this study proposes a residual-increment-oriented landslide displacement prediction framework that fuses multi-source monitoring variables. The displacement sequence is first processed into trend and periodic-related fluctuation representations, and the residual increment is used as the prediction target. Rainfall, reservoir water level, and the normalized difference vegetation index (NDVI) are incorporated as external monitoring variables. A cross-branch attention mechanism models interactions among heterogeneous feature branches, and a sparse MoE-based fusion module is introduced to adaptively adjust branch contributions under different deformation conditions. The model predicts the displacement residual increment, from which the final displacement is reconstructed. A case study using the Baishui River (Baishuihe) landslide monitoring dataset was conducted, together with additional validation on the related Bazimen Z110 landslide monitoring dataset and comparisons against conventional recurrent, convolutional, statistical, and Transformer-based baselines. The results show that the proposed model achieves lower RMSE and MAE than the compared methods on the tested datasets. These findings suggest that residual-increment modeling, multi-source monitoring variables, and condition-dependent branch fusion can improve short-term displacement prediction for the tested reservoir-area landslide cases.



Academic Editors: Vahdettin Demir and Çiğdem ACI

Received: 1 May 2026

Revised: 16 June 2026

Accepted: 22 June 2026

Published: 24 June 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

Keywords: landslide displacement prediction; multi-source monitoring variables; cross-branch attention; Mixture of Experts; residual-increment modeling; EWMA-based trend estimation; normalized difference vegetation index (NDVI)

1. Introduction

Landslides are major geological hazards that can threaten infrastructure, public safety, and human life [1]. Displacement prediction is an important component of landslide monitoring and risk assessment because it provides quantitative information for early warning

and decision-making [2]. With the development of deep learning, data-driven models have been increasingly used to extract temporal patterns from monitoring data and to integrate multi-source observations [3]. Recent studies have also shown that landslide monitoring and early warning can benefit from real-time displacement prediction models [4] and from the joint use of multi-source observations, such as satellite, hydrological, and ground-based monitoring data [5].

A common strategy for landslide displacement prediction is to decompose the cumulative displacement into components with different physical or statistical meanings, model these components separately, and then reconstruct the final displacement [6]. Previous studies have shown that hybrid models combining time-series analysis and deep learning can improve prediction accuracy [7]. Recurrent and convolutional neural networks, such as LSTM, GRU, and TCN, have also been applied to capture long-term dependencies and nonlinear deformation patterns in landslide displacement sequences [8–10]. In addition, residual-based analysis and prediction-interval methods have been used to describe regime shifts and uncertainty in landslide displacement forecasting [11,12]. These studies indicate that decomposition, residual modeling, and deep learning are useful tools for landslide displacement prediction; however, accurately modeling non-stationary cumulative displacement under changing environmental conditions remains challenging.

Temporal Convolutional Networks (TCNs) are widely used for time-series forecasting because dilated causal convolutions can capture long-range temporal dependencies with relatively efficient computation. Existing TCN-based forecasting methods have improved temporal representation learning for complex non-stationary sequences [13–20]. However, reservoir-area landslide deformation still exhibits evolving periodic responses and stage-dependent environmental effects that are difficult to represent using fixed temporal representations. Therefore, adaptive modeling of periodic features remains important for landslide displacement prediction.

Landslide displacement is affected not only by its historical deformation trend but also by external hydrological and environmental factors. Conventional multivariate models usually concatenate displacement, rainfall, reservoir water level, and other variables as early-fusion inputs [21,22]. More recent attention- and Transformer-based methods improve the interaction between endogenous displacement information and exogenous variables through attention or cross-attention mechanisms [23,24]. These methods provide useful tools for multi-source feature fusion, but they often treat the contributions of different information sources as relatively fixed. In reservoir-area landslides, the dominant factors may change across deformation stages, such as stable creep, rainfall-triggered acceleration, and reservoir-water-level-related fluctuation. Therefore, a fusion mechanism that can adaptively adjust the contributions of trend, periodic, and environmental representations may improve both prediction and interpretability. The Mixture of Experts (MoE) framework provides a divide-and-conquer strategy for condition-dependent representation learning, and related modular expert structures have been explored in geological-hazard tasks such as landslide susceptibility mapping [25]. However, the use of sparse routing for sample-level branch selection in landslide displacement prediction remains limited.

Existing landslide displacement prediction studies can be broadly grouped into three types. Decomposition-based hybrid models usually decompose the displacement sequence and predict each component separately. Conventional LSTM/GRU/TCN-based models often use early fusion of displacement and environmental variables. Recent attention- or Transformer-based models improve cross-source interaction through global attention or cross-attention. These approaches have advanced landslide displacement prediction, but they do not fully address the stage-dependent contributions of heterogeneous monitoring variables in reservoir-area landslides. Therefore, the novelty of this study does not lie in

any individual modeling component. Instead, it lies in a condition-dependent multi-source framework that jointly organizes trend, periodic-related fluctuation, and environmental representations through residual-increment prediction and adaptive branch fusion.

The main contributions of this study are as follows:

1. A residual-increment-oriented multi-source prediction framework is developed for reservoir-area landslide displacement. Unlike conventional decomposition-based frameworks that directly predict and superimpose decomposed displacement components, the proposed framework organizes trend, periodic-related fluctuation, and environmental information into separate branches, while NDVI is incorporated as vegetation-related auxiliary information. This formulation reduces the difficulty of directly modeling non-stationary cumulative displacement and links external triggering factors with short-term deformation changes.
2. A frequency-aware multi-scale enhanced TCN (MSE-TCN) module is designed for periodic-related fluctuation modeling. The module is embedded in the periodic branch and combines parallel causal dilated convolutions, channel recalibration, and frequency-domain conditional modulation to represent non-fixed, multi-scale periodic-related deformation patterns and vegetation-related seasonal variation.
3. A dual-level Attention–MoE fusion mechanism is proposed to separate time-step-level feature interaction from branch-level condition-dependent fusion. Cross-branch attention models interactions and possible phase shifts among trend, periodic, and environmental sequences, while sparse MoE routing adaptively adjusts the branch contributions for each sample and provides interpretable model-internal branch-weight indicators.

The proposed framework may provide a data-driven basis for offline reservoir-area landslide monitoring analysis and future short-term displacement forecasting or early-warning assessment for reservoir management authorities, geological-hazard monitoring agencies, emergency management departments, and infrastructure operators.

2. Materials and Methods

2.1. Overview of the Study Area

The Baishui River Landslide is located on the southern bank of the Yangtze River, approximately 56 km upstream of the Three Gorges Dam in Baishuihe Village, Shazhenxi Town, Zigui County. The landslide presents a stepped terrain with higher elevation in the south and lower elevation in the north. The elevation of the slope crest ranges from 450 m to 500 m. The terrain gradient is 24° – 36° within the elevation range of 180–500 m, while it is relatively gentle with a slope angle of 5° – 12° at elevations of 130–180 m, and reaches 27° – 31° at elevations of 80–130 m. In general, the front and rear edges of the landslide are steep, whereas the central part is relatively flat.

The exposed strata belong to the Lower Jurassic Xiangxi Formation (J_1x), which consists of yellowish-green and grayish-yellow siltstone, sandy shale and quartz sandstone interbedded with carbonaceous shale and coal seams. The stratum occurrence is 15° – 20° $\angle$ 30° – 35° , forming a bedding-controlled bank slope.

The landslide shows an irregular chair-shaped overall morphology. It is bounded by the Shanyang Gully on the west and Huangtubao hollow on the east. The elevation of the rear edge is about 297 m, and the front edge ranges from 120 m to 130 m, directly connecting to the Yangtze River. The landslide is about 500 m in longitudinal length and 430 m in width, covering an area of approximately $21.5 \times 10^4 \text{ m}^2$. The average thickness of the sliding mass is around 30 m, with a total volume of $645 \times 10^4 \text{ m}^3$. The main sliding direction is 20° .

A previous study has described the Baishui River Landslide as a typical large-scale colluvial reservoir-bank landslide in the Three Gorges Reservoir area [26]. According to the movement-type and material-type classification framework of Cruden and Varnes [27], the studied landslide can be classified as a large-scale deep-seated colluvial slide involving soil–debris material. This classification is supported by its irregular chair-shaped and stepped morphology, thick sliding mass, accumulation-layer material, and clear downslope movement direction. In terms of deformation mechanism, previous studies further indicate that the landslide exhibits traction-type progressive deformation and hydrodynamic-pressure-controlled deformation characteristics under reservoir-water-level drawdown and rainfall infiltration [28].

2.2. Relevant Dataset Description

The topographic map of the Baishui River Landslide and the continuous monitoring data spanning from January 2007 to December 2012 are presented in Figures 1 and 2. The dataset and topographic representation were sourced from the National Cryosphere Desert Data Center [29]. Unless otherwise stated, the main experiments in this study were conducted using the Baishui River ZG118 monitoring point.

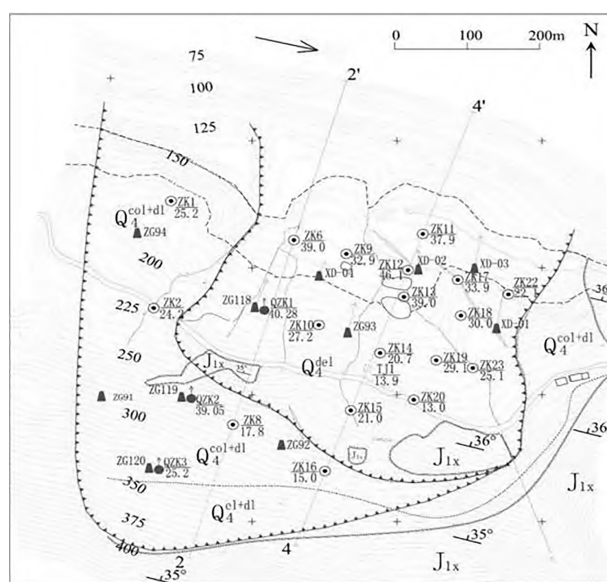


Figure 1. Topographic map of Baishui River Landslide.

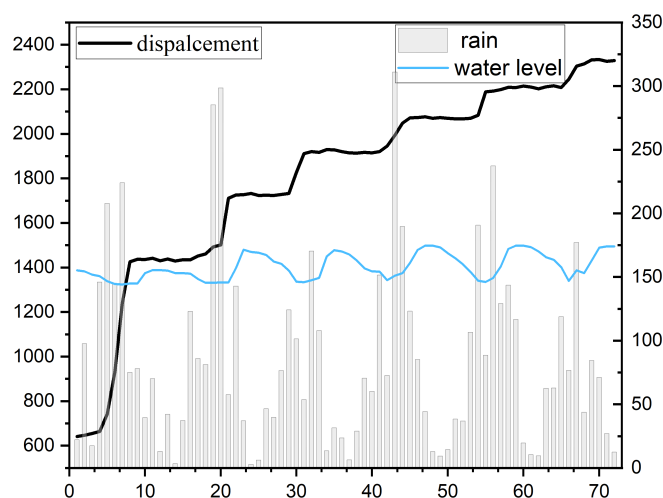


Figure 2. Monitoring data of the ZG118.

Vegetation cover is included as an auxiliary variable because it may reflect seasonal surface conditions related to slope stability, soil moisture, and erosion. In this study, vegetation information was represented by the normalized difference vegetation index (NDVI), which is calculated from near-infrared and red surface reflectance as follows:

$$\text{NDVI} = \frac{\text{NIR} - \text{RED}}{\text{NIR} + \text{RED}}, \quad (1)$$

where NIR and RED denote the surface reflectance in the near-infrared and red bands, respectively. NDVI ranges from -1 to $+1$, with larger values generally indicating denser or healthier vegetation cover. The NDVI data used in this study were obtained from the China 250 m monthly NDVI product covering the period from 2000 to 2022. This product was generated based on the Aqua/Terra-MODIS MOD13Q1 250 m 16-day NDVI product, the corresponding 250 m 16-day pixel reliability layer, and land-use data. In the product-generation procedure, MOD13Q1 data were first downloaded and preprocessed through format and projection conversion. Noise pixels identified by the pixel reliability layer were preliminarily reconstructed using neighboring pixels with the same land-use type within a 13×13 window. A long-term time-series Savitzky–Golay filter was then applied to further reconstruct the NDVI sequence, while original high-quality pixels were retained. Finally, the 16-day NDVI data were composited into monthly products using the maximum value composite method and mosaicked to generate the China-wide monthly NDVI dataset. The product has a spatial resolution of 250 m and a monthly temporal resolution, with the coordinate system of GCS_China_Geodetic_Coordinate_System_2000. The valid value range is from -2000 to 10000 , with a scale factor of 0.0001 . In this study, the monthly NDVI data from 2007 to 2012 were extracted to match the temporal coverage of the Baishui River landslide monitoring dataset. After applying the scale factor and excluding invalid or fill values, the mean NDVI value within the study-area ROI was calculated for each month and used as a vegetation-related auxiliary variable. The resulting monthly NDVI sequence was then aligned with the displacement, rainfall, and reservoir-water-level variables before subsequent temporal resampling and model input construction. Reliable landslide boundary vectors and precise georeferenced outlines are not available in the public monitoring dataset. Therefore, a boundary-overlaid NDVI map was not generated to avoid subjective spatial interpretation. Instead, NDVI was used as a temporally matched ROI-averaged auxiliary time-series input. GIS-based NDVI spatial visualization will be considered when verified boundary data become available.

2.3. Data Preprocessing

The multi-source monitoring variables were first organized at a consistent temporal scale before model training. The displacement and NDVI records were available at a monthly resolution, whereas rainfall and reservoir-water-level observations were recorded daily. The daily reservoir-water-level series contained no missing observations. For rainfall, the date sequence was continuous, but some rainfall-value cells were blank; these entries were treated as non-rainfall days and assigned a value of 0 mm before aggregation. The daily rainfall and reservoir-water-level records were then aggregated to the monthly scale and aligned with the monthly displacement and NDVI series. Occasional gaps that could arise during multi-source merging were handled using forward and backward filling.

To construct a sequence suitable for deep-learning-based temporal modeling under sparse monitoring conditions, the aligned monthly dataset was resampled to weekly resolution using Piecewise Cubic Hermite Interpolating Polynomial (PCHIP) interpolation. The weekly scale was selected as a compromise between the monthly displacement/NDVI records and the daily hydrological observations. This setting allows sub-monthly hydrolog-

ical variations to be represented while avoiding excessive daily interpolation of the monthly displacement series. Because the available displacement and NDVI records were monthly and limited in length, the interpolated weekly values were used as a reconstructed continuous temporal representation for model training, rather than as newly observed independent weekly measurements. Therefore, the PCHIP resampling step should be interpreted as an offline retrospective data-organization procedure. Because the reconstructed weekly values were derived from the original monthly observations, the resampling procedure increased the temporal density of the input sequence but did not increase the number of independently observed displacement records.

After the reconstructed weekly sequence had been obtained, the dataset was split chronologically into training and testing subsets. All normalization parameters were estimated from the training subset and then applied to the testing subset. The PCA transformation parameters used in the periodic branch were also fitted on the training subset and then applied to the testing subset. Thus, normalization, dimensionality reduction, and model fitting did not use test-set statistics.

In the reported experiments, the held-out testing subset was shorter than two complete seasonal cycles under the weekly resampling setting. Therefore, the one-sided exponentially weighted moving-average (EWMA) fallback was used to estimate the trend component for both the training and testing subsets. This EWMA-based trend estimation uses only the current and previous observations within each subset. The periodic-related fluctuation component was obtained by subtracting the estimated trend from the displacement sequence.

Based on the above preprocessing procedure, the present task is formulated as retrospective one-step-ahead prediction on a reconstructed weekly displacement sequence, rather than as a strict rolling-origin real-time deployment protocol. The input sequence length was set to 26 time steps. Other model architecture, optimization, regularization, and runtime settings are provided in Appendix A.

Figure 3 illustrates the landslide displacement prediction method proposed in this paper.

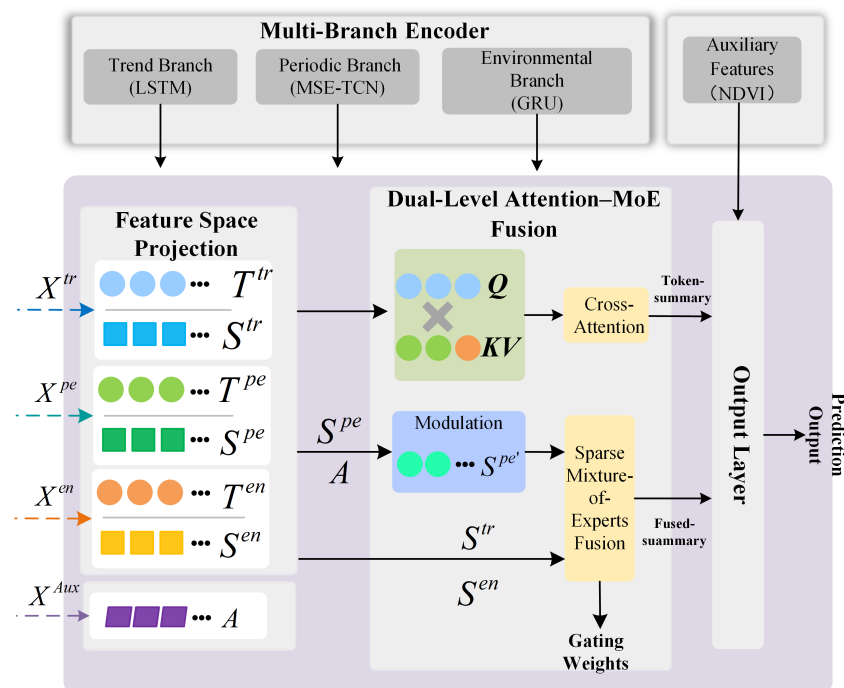


Figure 3. Workflow of the proposed framework.

2.4. Multi-Source Feature Extraction

(1) Trend branch: The trend representation is estimated from the displacement sequence using the EWMA-based trend-estimation strategy described in Section 2.3. It represents the long-term deformation tendency of the landslide and provides a low-frequency displacement representation for subsequent modeling.

(2) Environmental branch: Rainfall and reservoir-water-level variables are used as environmental inputs. These variables describe hydrological forcing conditions that may influence landslide deformation through rainfall infiltration, groundwater adjustment, and reservoir-bank slope loading or unloading.

(3) Periodic branch and NDVI auxiliary variable: The detrended fluctuation component represents periodic-related deformation variation. It is combined with NDVI and transformed using PCA to form the periodic-branch input. In addition, NDVI is also included in the auxiliary input together with total displacement to provide vegetation-related seasonal context for the fusion module. This design allows the model to use NDVI both as periodic-related information and as auxiliary contextual information.

2.5. MSE-TCN Module

Conventional Temporal Convolutional Networks (TCNs) usually use single-scale dilated convolutions to model temporal dependencies. However, landslide displacement sequences may contain multi-scale oscillations, variable time lags, and overlapping frequency components. A single-scale structure may therefore be insufficient for representing these non-stationary patterns. To better represent evolving periodic deformation patterns, the periodic branch is enhanced using multi-scale temporal feature extraction and adaptive modulation operations, as shown in Figure 4.

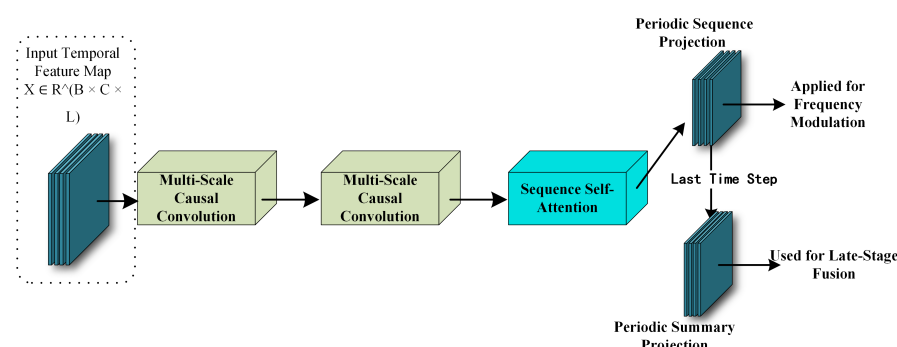


Figure 4. Structure of the MSE-TCN module.

2.6. Multi-Scale Parallel Dilated Convolution

Different temporal patterns may occur at different time scales. Therefore, multiple convolution branches are used to capture short- and long-range periodic responses simultaneously. Let the periodic branch input sequence be $X \in \mathbb{R}^{L \times C}$, where L denotes the sequence length and C the number of channels. MSE-TCN employs j parallel dilated convolution branches, each configured with a distinct kernel size and dilation rate, to extract local patterns characterized by varying time constants:

$$H_i = \phi(\text{Conv}_{k_i, d_i}(X)), \quad i = 1, 2, \dots, j, \quad (2)$$

where $\phi(\cdot)$ denotes nonlinear activation and regularization operations. Subsequently, the multi-scale outputs are concatenated channel-wise and fused back into the same dimension using convolution:

$$H = \text{Conv}(\text{Concat}(H_1, H_2, \dots, H_j)). \quad (3)$$

The causal convolutions of the original TCN are retained so that the convolutional module uses only previous time steps within each input sequence. The multi-scale parallel convolutional structure is illustrated in Figure 5.

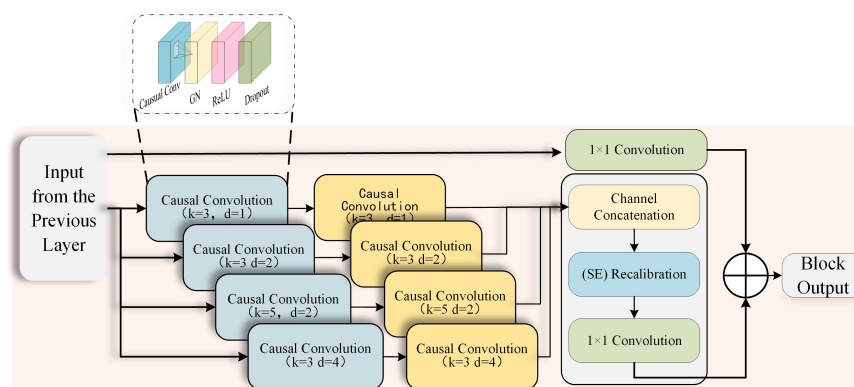


Figure 5. Multi-scale parallel causal convolutional block.

2.7. Channel Recalibration

Multi-scale parallel structures introduce additional scale channels; however, the effectiveness of these channels varies across different stages. To enable adaptive channel selection, Squeeze-and-Excitation (SE) recalibration is introduced. First, global average pooling is applied to H to obtain channel descriptors:

$$z_c = \frac{1}{L} \sum_{t=1}^L H_{t,c}. \quad (4)$$

Then, channel weights are generated through a gating network:

$$s = \sigma(W_2 \delta(W_1 z)), \quad (5)$$

where $\delta(\cdot)$ and $\sigma(\cdot)$ denote the ReLU and sigmoid functions, respectively. The recalibrated output is

$$\tilde{H}_{t,c} = s_c \cdot H_{t,c}. \quad (6)$$

Finally, a residual connection is added to form a stable gradient path.

2.8. Frequency-Domain Conditional Modulation

Landslide deformation may exhibit evolving periodic responses under changing hydrological and environmental conditions.

Specifically, let the input sequence be $x(t)$. The short-time Fourier transform is first applied to obtain its frequency-domain representation:

$$X(\tau, \omega) = \sum_{t=-\infty}^{+\infty} x(t) w(t - \tau) e^{-j\omega t}, \quad (7)$$

where $w(\cdot)$ denotes the window function.

The power spectrum is then computed and averaged over time to obtain the spectral energy distribution:

$$P(\omega) = \frac{1}{T} \sum_{\tau=1}^T |X(\tau, \omega)|^2. \quad (8)$$

Subsequently, the frequency axis is partitioned into several bands and pooled to obtain the frequency-band representation g_f . This representation is combined with the auxiliary

context embedding (A), which is generated from the auxiliary input, and then fed into a lightweight MLP to generate the modulation coefficient:

$$\alpha = \text{MLP}([g_f; A]). \quad (9)$$

The modulation coefficient is then used to adjust the periodic-branch representation:

$$\hat{H} = \alpha \odot \tilde{H}. \quad (10)$$

In the proposed module, the periodic-branch representation is first transformed into a compact frequency-domain descriptor. This descriptor is combined with the auxiliary context embedding and used to adaptively modulate the periodic representation according to the current spectral and auxiliary conditions.

2.9. Dual-Level Attention–MoE Fusion (DA–MoE Fusion)

Landslide displacement prediction requires the integration of heterogeneous monitoring variables with different temporal responses. Trend, periodic, and environmental information may contribute differently under stable creep, rainfall-triggered acceleration, and reservoir-water-level-related fluctuation stages. To model these stage-dependent contributions, this study introduces a dual-level Attention–MoE fusion module.

Cross-branch attention first captures temporal interactions among trend, periodic, and environmental representations. In the implemented model, the trend-branch tokens are used as queries, while the concatenated periodic-branch and environmental-branch tokens are used as keys and values:

$$h_a = \text{Pool}(\text{CrossAttn}(Q_t, K_{pe}, V_{pe})) \quad (11)$$

where Q_t denotes the projected trend-branch tokens, and K_{pe} , V_{pe} are obtained from the concatenated periodic and environmental token representations.

Branch-level representations are then adaptively weighted to reflect different deformation conditions:

$$h_m = \sum_{i=1}^3 \gamma_i h_i, \quad \sum_{i=1}^3 \gamma_i = 1, \quad (12)$$

where γ_i is generated by a Top- k sparse gating network and h_i represents the summary of each branch.

Finally, the outputs from the two stages are concatenated and mapped into a unified fused representation:

$$h_f = \text{MLP}([h_a; h_m]). \quad (13)$$

Although only three branch representations are used in this study, the sparse routing mechanism is introduced as a sample-level condition-dependent branch selection strategy rather than primarily for computational acceleration. In reservoir-area landslide displacement prediction, the dominant information source may vary across deformation stages, such as stable creep, rainfall-triggered acceleration, and reservoir-water-level-induced fluctuation. Compared with dense weighted fusion, Top- k sparse routing activates the most relevant branch representations for each sample and provides model-internal indicators of condition-dependent branch contributions. In the current framework, sparse routing is therefore used mainly to enhance dynamic branch interpretability. A more systematic comparison between sparse and dense fusion strategies will be examined in future work.

2.10. Displacement Prediction

To address this issue, this study uses a residual-increment prediction strategy. The displacement increment is defined as

$$\Delta y_t = y_t - y_{t-1}. \quad (14)$$

The model learns the mapping from historical multi-source feature sequences within the look-back window to the displacement increment at the next time step. During evaluation, the predicted total displacement is obtained through one-step reconstruction:

$$\hat{y}_t = y_{t-1} + \widehat{\Delta y}_t. \quad (15)$$

In this study, historical multi-source observations within the look-back window are used to predict the displacement residual increment at the next time step. The final displacement prediction is then obtained through one-step reconstruction according to Equation (15). Considering the limited size of long-term landslide monitoring datasets, the current experiments focus on next-step forecasting to avoid cumulative error propagation in recursive multi-step prediction. Under the weekly resampling setting used in this study, the next-step prediction corresponds to the next step of the reconstructed weekly displacement sequence. This retrospective prediction setting may provide a basis for future short-term displacement monitoring and early-warning studies when longer and more frequently updated monitoring records become available.

2.11. Evaluation Metrics

To evaluate model performance and further analyze prediction accuracy, this paper uses three evaluation metrics: root mean square error (RMSE), coefficient of determination (R^2), and mean absolute error (MAE). The calculation formulas are as follows:

$$\text{RMSE} = \sqrt{\frac{1}{m} \sum_{i=1}^m (y_i - \hat{y}_i)^2}, \quad (16)$$

$$R^2 = 1 - \frac{\sum_{i=1}^m (y_i - \hat{y}_i)^2}{\sum_{i=1}^m (y_i - \bar{y})^2}, \quad (17)$$

$$\text{MAE} = \frac{1}{m} \sum_{i=1}^m |y_i - \hat{y}_i|, \quad (18)$$

where y_i is the actual value at time i , $\hat{y}_i$ is the predicted value at time i , $\bar{y}$ is the mean of the measured data, and m is the sample size.

3. Results

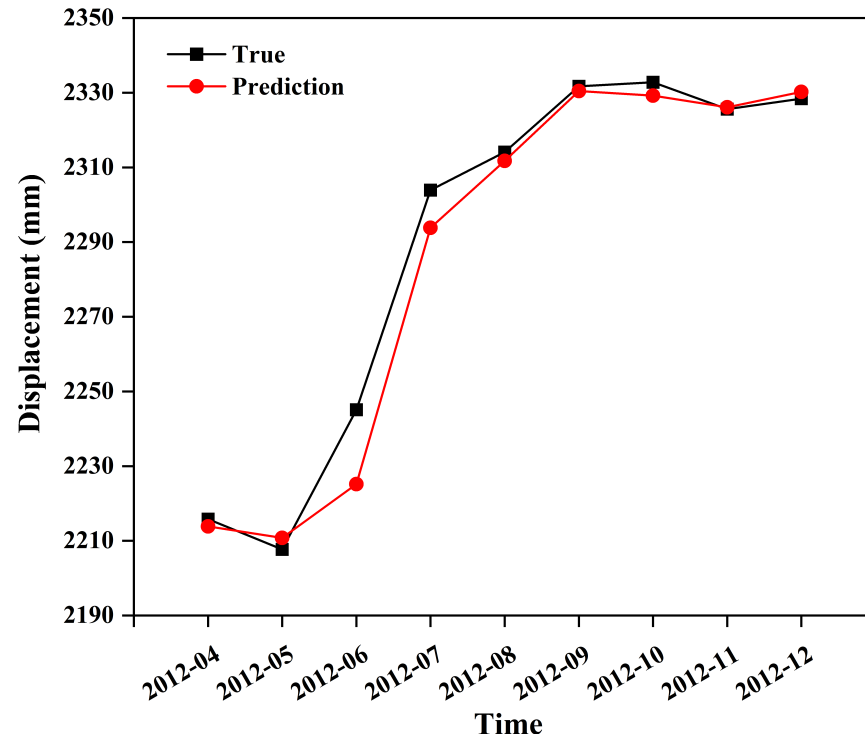
3.1. Comparative Experiment Setup

To provide a comprehensive comparison, both conventional forecasting models and recent attention-based time-series models were selected as baselines. The conventional baselines included LSTM, GRU, ARIMA, while the attention-based baselines included Transformer, Informer-style, and Autoformer-style models. For a fair comparison, all trainable neural-network baselines were evaluated using the same preprocessed weekly dataset, chronological training/testing split, look-back window, and available input variables as the proposed model. The ARIMA model was implemented as a univariate statistical baseline using the displacement sequence, since it does not directly support the same multivariate input structure. Results are presented in Table 1.

Table 1. Summary of comparative experimental results.

Model	RMSE (mm)	R^2	MAE (mm)
LSTM	11.3225	0.9238	8.6713
GRU	10.3418	0.9348	7.2146
ARIMA	12.5715	0.9150	9.2624
Transformer	4.3443	0.9916	3.1137
Informer-style	5.0697	0.9885	3.5904
Autoformer-style	4.3392	0.9914	2.7676
Proposed model	3.3789	0.9949	2.4473

As shown in Table 1, the proposed model achieved the lowest RMSE and MAE and the highest R^2 among the compared methods. Compared with the conventional recurrent, convolutional, statistical, and attention-based baselines, the proposed model reduced prediction errors on the tested dataset. These descriptive results suggest that residual-increment prediction combined with multi-branch representation learning, token-level cross-branch attention, and sparse MoE fusion can improve short-term displacement prediction under the same preprocessing and evaluation protocol. The selected baselines cover both commonly used landslide displacement prediction model families and recent deep time-series forecasting architectures. Because geohazard-specific models reported in previous studies often differ in monitoring sites, sampling intervals, input variables, prediction horizons, and preprocessing strategies, this study emphasizes controlled within-dataset comparisons rather than direct cross-paper numerical ranking. The comparison between the measured and predicted displacement during the held-out testing period is shown in Figure 6.

**Figure 6.** Comparison of measured and predicted displacement during the held-out testing period.

Among the compared baselines, Autoformer-style achieved the lowest RMSE and MAE, whereas Transformer achieved the highest R^2 . Relative to Autoformer-style, the proposed model reduced RMSE from 4.3392 mm to 3.3789 mm and MAE from 2.7676 mm

to 2.4473 mm, corresponding to relative reductions of 22.13% and 11.57%, respectively. Compared with Transformer, the R^2 increased from 0.9916 to 0.9949. These differences are reported as descriptive measures of performance improvement rather than as evidence of formal statistical significance. Because the held-out period contains only a limited number of original monthly displacement observations, the reconstructed weekly testing values are temporally correlated and cannot be regarded as additional independent observations. Treating all reconstructed weekly points as independent samples in a formal significance test would therefore overstate the effective test-set size. Accordingly, the results are interpreted as descriptive evidence of improved performance under the current evaluation protocol rather than as statistically confirmed superiority.

3.2. Computational Complexity Analysis

To evaluate the computational complexity and efficiency of the proposed model, we compared it with several baseline architectures, including Simple LSTM, Transformer, Autoformer-style, and Informer-style networks. All models were trained using the same preprocessed weekly displacement dataset with identical training/test splits, look-back window, and input variables. Training was performed on a CPU environment, and the reported metrics include model parameter count, model size, average training time per epoch, total epoch time, and inference time per sample. These settings ensure a fair comparison of computational cost while maintaining consistent data and training configurations across models. The corresponding computational complexity and efficiency results are summarized in Table 2.

Table 2. Computational complexity and efficiency comparison.

Model	Params (M)	Model Size (MB)	Training Time (s)	Epoch Time (s)	Inference Time (ms/Sample)
Simple LSTM	0.027	0.102	3.35 ± 0.62	0.037 ± 0.007	0.0393 ± 0.0023
Transformer	0.424	2.104	15.94 ± 4.25	0.157 ± 0.004	0.1108 ± 0.0108
Autoformer-style	1.473	3.694	21.43 ± 2.06	0.261 ± 0.018	0.1222 ± 0.0099
Informer-style	0.873	2.784	16.12 ± 3.07	0.169 ± 0.004	0.1489 ± 0.0038
Proposed model	2.518	9.622	83.99 ± 15.00	0.695 ± 0.014	0.5450 ± 0.0154

3.3. External Validation and Repeated-Run Evaluation

To further evaluate prediction consistency beyond a single run and the main dataset, the Bazimen landslide Z110 monitoring point dataset was included as an additional validation case [30]. Because the Baishui River and Bazimen datasets represent related reservoir-area landslide monitoring contexts, the Bazimen dataset is used here to provide supplementary validation within a similar geographical and monitoring setting. For each dataset, five repeated experiments were conducted using the same five random seeds: 42, 2024, 2025, 3407, and 1234. These repeated runs were used only to evaluate run-to-run consistency, whereas the default comparative, ablation, visualization, and interpretability experiments were conducted using seed 42 unless otherwise stated. For this supplementary validation and repeated-run analysis, the mean and standard deviation of RMSE, MAE, and R^2 were reported to evaluate prediction accuracy and run-to-run variation.

As shown in Table 3, the proposed model showed relatively limited run-to-run variation on both the Baishui River and Bazimen Z110 datasets under the current repeated-seed evaluation protocol. Across the ten dataset-seed combinations, corresponding to two datasets and five seeds per dataset, the overall RMSE, MAE, and R^2 were 4.3716 ± 0.8488 mm, 2.9086 ± 0.4466 mm, and 0.9935 ± 0.0025 , respectively. These results

provide supplementary validation in a related reservoir-area monitoring context, while broader cross-domain generalization remains to be evaluated.

Table 3. External validation and repeated-run results.

Dataset	RMSE (mm)	R^2	MAE (mm)
Bazimen Z110	4.8172 ± 0.3032	0.9928 ± 0.0007	3.1818 ± 0.2261
Original landslide	3.9261 ± 1.0162	0.9943 ± 0.0034	2.6354 ± 0.4592
Overall	4.3716 ± 0.8488	0.9935 ± 0.0025	2.9086 ± 0.4466

To further assess performance consistency across chronological partitions and random initializations, two supplementary analyses were conducted: an expanding-window rolling-origin evaluation and a paired repeated-seed comparison with the TCN baseline. The rolling-origin evaluation used five chronological origins, corresponding to training fractions of 60%, 65%, 70%, 75%, and 80%, and was conducted using the default seed 42. Therefore, this evaluation was used to assess performance variation across different chronological train/test partitions rather than variation across repeated random seeds. The averaged results across the five rolling origins, obtained using seed 42, yielded an RMSE of 5.5389 mm, an MAE of 3.327 mm, and an R^2 of 0.9818. This result indicates that the model maintained reasonable one-step-ahead performance under different chronological partitions.

In the paired repeated-seed comparison, the proposed model and the TCN baseline were trained and evaluated under the same chronological split, input variables, residual-increment prediction target, and preprocessing protocol. The same five seeds used in the repeated-run evaluation, namely 42, 2024, 2025, 3407, and 1234, were adopted for this paired comparison. Compared with the TCN baseline, the proposed model reduced RMSE from 3.5880 ± 0.2126 mm to 3.4734 ± 0.4572 mm and MAE from 2.8071 ± 0.2421 mm to 2.6008 ± 0.2569 mm. The R^2 also increased from 0.9941 ± 0.0007 to 0.9954 ± 0.0015 . These results suggest that the reported improvement is not solely dependent on a single train/test split or random initialization.

3.4. Lightweight Uncertainty Assessment

To provide a preliminary indication of predictive uncertainty, a lightweight ensemble-based analysis was conducted. Five models were trained using the same five random seeds as the repeated-run stability evaluation, namely 42, 2024, 2025, 3407, and 1234, and their predictions were aggregated to estimate an empirical prediction range. On the monthly evaluation scale, the ensemble mean generally followed the measured displacement trend. The mean predictive standard deviation was 4.8113 mm, and the average width of the 95% empirical prediction interval was 11.1551 mm. As shown in Figure 7, the shaded interval provides an intuitive indication of predictive uncertainty around the ensemble mean prediction. This interval should be interpreted as a lightweight empirical uncertainty indicator rather than a formally calibrated probabilistic prediction interval. More rigorous uncertainty calibration methods will be considered in future work.

3.5. Ablation Study Setup

To evaluate the contribution of different components of the proposed framework, several ablation experiments were conducted. The results are shown in Table 4. As shown in Table 4, the proposed model yielded the most favorable combination of RMSE and MAE and R^2 among the tested variants. Removing the cross-attention mechanism, the MoE-based fusion module, the auxiliary NDVI information, or the SE module in the MSE-TCN branch all degraded the prediction performance to different extents. In particular, removing the SE module increased the RMSE to 10.3621 mm and the MAE to 6.1777 mm,

while reducing R^2 to 0.9513, indicating that channel-wise feature recalibration contributes to the representation of informative temporal features in the periodic branch.

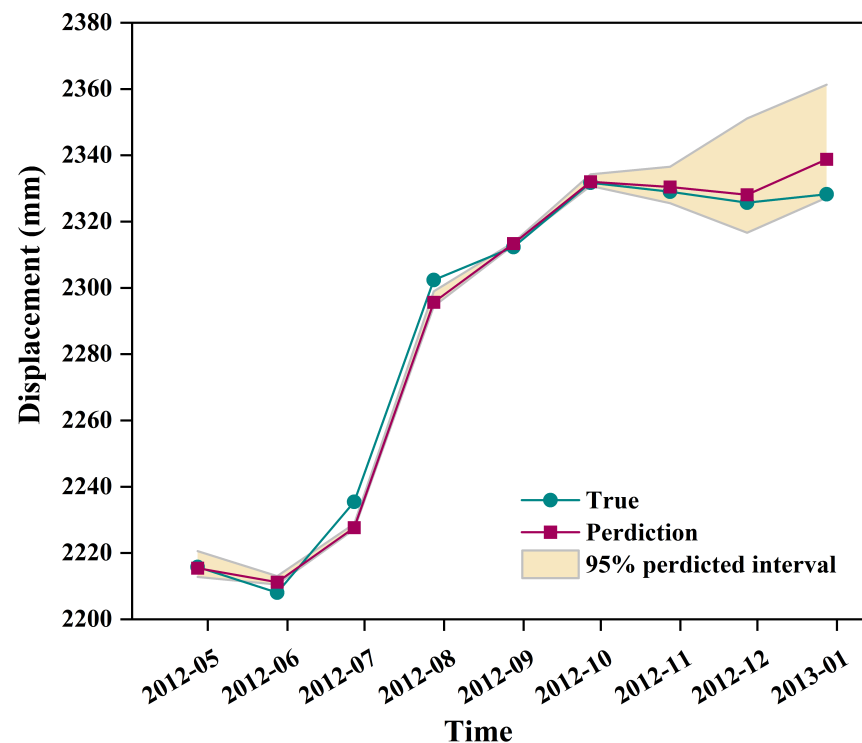


Figure 7. Ensemble-based empirical prediction interval for monthly landslide displacement.

Because the ablation experiments were conducted as separate controlled ablation batches, Table 4 is used to compare the relative performance changes caused by removing individual components under the same experimental setting. Compared with previous decomposition-based landslide displacement prediction methods, which usually decompose cumulative displacement and predict each component separately, the present framework focuses on residual-increment prediction after multi-source feature organization. This reduces the difficulty of directly modeling highly non-stationary cumulative displacement.

Table 4. Ablation experiment results.

Experimental Setting	RMSE (mm)	R^2	MAE (mm)
Direct fusion without Cross-attention	3.6121	0.9844	3.0866
Without MoE	16.7498	0.8727	8.0590
Without NDVI auxiliary variable	13.4726	0.8742	9.4498
Without SE in MSE-TCN	10.3621	0.9513	6.1777
Proposed model	3.3789	0.9949	2.4473

4. Discussion

4.1. Limitations and Applicability of Methods

Although the proposed framework showed favorable performance under the current evaluation protocol, several limitations should still be acknowledged. The current results are most applicable to retrospective or offline reservoir-area landslide monitoring analyses in which sparse displacement records are integrated with hydrological and vegetation-related variables. Broader validation on additional landslide cases is still required before stronger generalization claims can be made, because geological settings and triggering mechanisms vary substantially among landslides. The Bazimen Z110 evaluation extends

the analysis to an additional related landslide dataset, but it does not constitute validation across a substantially different geological or monitoring domain. Potential overfitting risks may also remain, despite the use of repeated experiments, regularization strategies, chronological train/test splitting, and training-set-only scaling. This limitation is largely associated with the limited size of publicly available long-term landslide monitoring datasets.

More specifically, the held-out test set contains only a limited number of original monthly displacement observations. Although PCHIP resampling increases the temporal density of the sequence, it does not create additional independently observed displacement measurements, because the reconstructed weekly values are derived from the same underlying monthly records. In addition, the testing period covers fewer than two complete seasonal cycles under the reconstructed weekly setting. Consequently, the reported performance metrics may not fully represent the variability associated with a broader range of deformation stages, seasonal transitions, and hydrological conditions. The repeated-seed and rolling-origin evaluations provide supplementary evidence of run-to-run and temporal-partition consistency under the current protocol, but they cannot substitute for validation using longer and independently observed displacement records. Therefore, the present results should be interpreted within the tested retrospective setting rather than as conclusive evidence of broad temporal generalizability or operational robustness.

The same limitation also constrains formal statistical significance analysis. Because the reconstructed weekly testing values are temporally correlated and originate from the same underlying monthly observations, treating them as independent paired samples could overstate the effective test-set size and produce misleading significance estimates. Moreover, repeated runs with different random seeds characterize variability associated with model initialization and optimization rather than independent replication of the monitoring observations. Accordingly, no formal statistical significance claim is made, and the performance differences relative to the strongest baselines are reported descriptively under the current evaluation protocol. In addition, the framework relies on temporal aggregation, interpolation, EWMA-based trend estimation, and normalization; its performance may therefore be sensitive to preprocessing configurations and sampling resolution.

To further evaluate the robustness of the proposed model under imperfect monitoring conditions, a lightweight hydrological-input perturbation test was conducted on the held-out test set. The trained model was evaluated under four perturbed settings, including 10% missing hydrological inputs with training-mean imputation, 20% missing hydrological inputs with training-mean imputation, a one-week hydrological-input delay, and a combined 10% missing-input plus one-week delay setting. Compared with the clean test setting, hydrological-input masking caused only minor changes in prediction accuracy, with absolute relative changes of 4.14% and 5.04% in RMSE and 2.61% and 3.82% in MAE under the 10% and 20% missing-input settings, respectively. The one-week delayed-update scenario also resulted in limited changes, with absolute relative changes of 3.03% in RMSE and 0.95% in MAE. Under the combined missing-input and delay setting, the absolute relative changes in RMSE and MAE were 2.94% and 0.95%, respectively. These results indicate that the proposed model maintained comparable prediction accuracy under moderate missing or delayed hydrological inputs, suggesting that the model is not overly sensitive to such limited monitoring imperfections in this test case. Therefore, this experiment provides a useful supplementary robustness check, although it should not be interpreted as a comprehensive operational missing-data evaluation.

However, long-term sensor failure, systematic data latency, automatic data updating, and uncertainty-aware warning strategies still require further investigation before operational early-warning applications. In particular, the current uncertainty assessment is based on an empirical ensemble prediction range rather than a formally calibrated probabilistic

prediction interval. It should therefore be interpreted as a lightweight uncertainty indicator rather than a fully calibrated safety-critical uncertainty estimate. Future work should examine calibrated uncertainty quantification methods, such as conformal prediction, quantile regression, or Bayesian neural networks. Hybrid approaches that integrate physically based geotechnical mechanisms with deep learning models should also be explored to improve interpretability and predictive robustness.

4.2. Interpretability Analysis

To examine the interpretability of the fusion module, the dynamic MoE branch weights were analyzed, as shown in Figure 8. The figure visualizes how the model assigned different weights to the trend, periodic, and environmental branches during the testing period. The weights varied over time, indicating that the model did not use fixed contributions from different information sources. The periodic branch maintained relatively high weights in most testing months, suggesting that periodic-related fluctuation information provided a stable contribution to prediction. The trend branch became more prominent in several months, while the environmental branch showed intermittent increases, indicating that the model adjusted the relative importance of displacement-history and hydrological information under different deformation conditions. This visualization illustrates the condition-dependent branch selection behavior of the fusion module. However, these branch weights should be interpreted as model-internal contribution indicators rather than direct causal evidence of landslide deformation mechanisms.

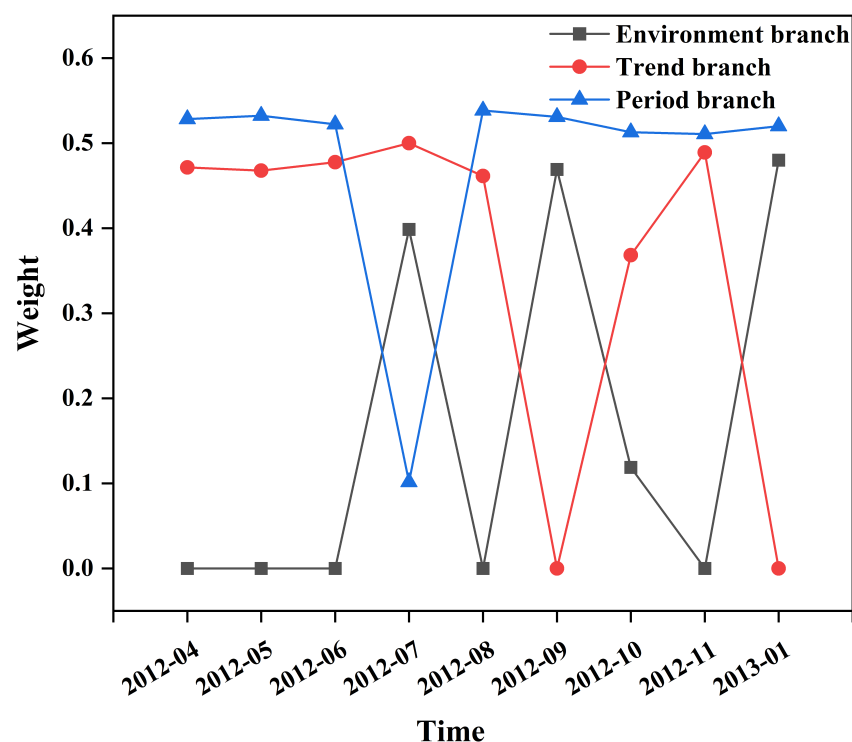


Figure 8. Dynamic MoE branch weights during the testing period.

To further verify the contribution of NDVI beyond qualitative physical interpretation, we conducted two lightweight analyses: NDVI perturbation sensitivity analysis and permutation importance analysis. In the perturbation analysis, the standardized NDVI input was changed by -1.0 , -0.5 , 0 , $+0.5$, and $+1.0$ standard deviation units, while the other inputs remained unchanged. In the permutation importance analysis, the NDVI feature was randomly shuffled 10 times, and the increase in total-displacement RMSE was used to

quantify its importance. In addition, the lag relationship between NDVI and displacement increment was examined, and the best lag was found to be 0 weeks in the current dataset.

As shown in Table 5, the sensitivity and permutation analyses indicate that NDVI provided useful predictive information in the tested case. Decreasing NDVI by 1.0 standardized unit increased the RMSE from 3.3789 mm to 5.7823 mm, while NDVI permutation increased the RMSE by 0.9233 ± 0.2470 mm on average over 10 repetitions. The lag analysis showed that the strongest relationship with displacement increment occurred at 0 weeks, suggesting a mainly contemporaneous NDVI–deformation relationship in this dataset.

In addition to the NDVI-related analysis, a post hoc hydrological event-response analysis was conducted for the Baishui River Landslide. As shown in Table 6, rainfall showed the strongest response at a lag of 3 weeks, with a main response window of 1–7 weeks after rainfall events, whereas the rainfall-event intensity indicator showed its strongest response at a lag of 6 weeks. This suggests a delayed rainfall–deformation response, which is consistent with rainfall infiltration and groundwater adjustment processes. The MoE weights further indicate that all three branches contributed to the prediction, with larger mean weights for the trend and periodic branches and a smaller but non-negligible contribution from the environmental branch.

Table 5. NDVI sensitivity and permutation importance analysis results.

Analysis Type	Setting	Evaluation Metric	Result
Baseline prediction	Original NDVI	RMSE/MAE/ R^2	3.3789/2.4473/0.9949
NDVI perturbation	−1.0 std	RMSE/MAE/ R^2	5.7823/3.2601/0.9898
NDVI perturbation	−0.5 std	RMSE/MAE/ R^2	4.2105/2.9284/0.9921
NDVI perturbation	+0.5 std	RMSE/MAE/ R^2	2.4300/1.8961/0.9974
NDVI perturbation	+1.0 std	RMSE/MAE/ R^2	2.1751/1.6660/0.9929
NDVI permutation importance	10 repeats	Δ RMSE	0.9233 ± 0.2470 mm
NDVI permutation importance	10 repeats	Δ RMSE (%)	$27.33 \pm 7.31\%$
NDVI lag analysis	Displacement increment	Best lag	0 weeks

Table 6. Hydrological response and environmental-branch interpretability for the Baishui River Landslide.

Analysis Type	Variable/Branch	Indicator	Result
Hydrological event response	Rainfall	Strongest lag	3 weeks
Hydrological event response	Rainfall	Mean response at strongest lag	24.6374
Hydrological event response	Rainfall	Main response window	1–7 weeks
Hydrological event response	Maximum rainfall	Strongest lag	6 weeks
MoE branch weight	Environmental branch	Mean weight	0.1337
MoE branch weight	Trend branch	Mean weight	0.4013
MoE branch weight	Periodic branch	Mean weight	0.4650

The learned branch behaviors are broadly consistent with the deformation characteristics expected for reservoir-area landslides. The mean MoE weights show that the model did not rely on a single information source. Instead, the relatively large contributions of the trend and periodic branches suggest that historical displacement evolution and seasonal variation were important for the tested prediction task. The environmental branch also contributed to the prediction, in agreement with the hydrological response analysis, which showed that rainfall-related indicators were associated with displacement changes at specific time lags. Such delayed responses are commonly linked to rainfall infiltration and hydrological adjustment in slopes. Reservoir-water-level fluctuation may further modify the hydraulic boundary conditions of reservoir-bank slopes and interact with rainfall-related processes. NDVI provides auxiliary seasonal information related to vegetation cover and slope-surface conditions. These results should therefore be interpreted as physically plausible model-internal evidence, rather than as direct causal attribution.

4.3. Interpretation of Key Experimental Results

The proposed model has higher computational complexity than the simpler baseline architectures. It contains 2.518 million parameters and requires approximately 0.545 ms per sample for inference. This increased cost is primarily due to the multi-branch encoder, MSE-TCN module, token-level cross-branch attention, and sparse MoE fusion. Nevertheless, the overall model size remains below 10 MB and the inference time is less than 1 ms per sample, indicating that the proposed approach remains computationally feasible for offline landslide displacement prediction and monitoring applications. In addition to computational feasibility, the results can be further interpreted in relation to recent landslide displacement prediction studies.

Recent landslide displacement prediction studies have also shown that decomposition-based preprocessing, temporal deep-learning models, and multi-source environmental variables are effective for modeling nonlinear landslide deformation. Lin et al. [31] combined ICEEMDAN decomposition with a TCN–BiLSTM network and showed that separating trend and fluctuation components helps improve landslide displacement prediction under the influence of rainfall, groundwater level, and historical displacement. Li et al. [32] used an Informer-based model with InSAR-derived deformation and environmental variables to predict reservoir-bank slope deformation, highlighting the value of Transformer-style architectures for long-sequence deformation forecasting. Chen et al. [33] proposed a TCN–Multihead-Attention model for landslide deformation prediction in the Three Gorges Reservoir area and showed that combining causal temporal convolutions with attention mechanisms can improve the representation of deformation-related temporal interactions. These studies are consistent with the present results, which indicate that landslide displacement prediction benefits from temporal feature organization, multi-source feature integration, and attention-based interaction modeling.

Compared with these recent studies, the proposed framework has several distinctive characteristics. First, instead of predicting decomposed displacement components and directly reconstructing cumulative displacement, the present model predicts residual increments on a reconstructed weekly displacement sequence, which reduces the difficulty of directly modeling highly non-stationary cumulative displacement. Second, rather than treating all monitoring variables as a single early-fusion input, the proposed framework organizes trend, periodic-related fluctuation, and environmental information into separate branches, while NDVI is incorporated as vegetation-related auxiliary information for contextual fusion. Third, the sparse MoE module provides sample-level, condition-dependent branch fusion, which helps describe how the relative contributions of trend, periodic, and environmental representations vary during different deformation stages. Therefore, the contribution of this study does not lie in the isolated use of TCN, attention, temporal feature processing, or MoE. Instead, it lies in integrating residual-increment modeling, physically motivated multi-source branches, cross-branch attention, and sparse condition-dependent fusion for the two tested reservoir-area landslide cases.

The ablation experiments further clarify the contributions of each module within the proposed framework.

The removal of cross-attention indicates that time-step-level interactions among heterogeneous branches are important for aligning trend, periodic, and environmental information. Similarly, the absence of the MoE-based fusion module shows that adaptive combination of branch representations contributes to capturing condition-dependent deformation patterns. Excluding NDVI reduces predictive performance, confirming the value of vegetation-related auxiliary information. The single-branch TCN variant also performed worse than the multi-branch model, highlighting the importance of a structured multi-branch representation for organizing heterogeneous landslide monitoring variables.

4.4. Practical Implications and Future Work

The proposed framework may support offline reservoir-area landslide monitoring analysis and provide a data-driven basis for future short-term displacement forecasting and early-warning assessment. The model dynamically adjusts the contributions of different information sources according to landslide deformation stages. The learned branch weights suggest that the model adaptively combines displacement-history, seasonal, and environmental information under different deformation conditions. NDVI provides seasonal context related to vegetation cover and slope-surface conditions. By predicting displacement residual increments rather than absolute displacement, the model reduces the difficulty of modeling cumulative displacement and helps identify short-term deformation changes.

The retrospective setting adopted in this study is mainly constrained by the limited length and monthly resolution of the available displacement and NDVI records. Under such data conditions, weekly reconstruction and trend estimation provide a practical way to organize sparse and asynchronous monitoring variables for sequence-model training. With denser, longer, and more frequently updated monitoring records, the proposed multi-branch Attention–MoE framework could be further extended, in future work, to rolling-origin or near-real-time prediction by updating the temporal reconstruction, trend-estimation, and model-input construction procedures at each prediction origin.

Although the repeated-run results showed relatively limited run-to-run variation on the Baishui River and related Bazimen datasets under the current evaluation protocol, the two datasets share a similar reservoir-area monitoring context. Broader generalization across substantially different geological conditions, geographical regions, and monitoring systems therefore remains to be evaluated. Future work should evaluate additional landslide cases with different geological conditions, extend the framework to multi-step or rolling-origin forecasting, and investigate hybrid approaches that integrate physically based geotechnical mechanisms with data-driven models.

5. Conclusions

This study developed a residual-increment-oriented framework for reservoir-area landslide displacement prediction by integrating trend, periodic-related fluctuation, and environmental representations through a dual-level Attention–MoE fusion module, with NDVI incorporated as vegetation-related auxiliary information. The framework was evaluated on the Baishui River Landslide dataset and compared with conventional recurrent, convolutional, statistical, and Transformer-based baselines.

The results show that the proposed framework achieved lower RMSE and MAE than the compared baseline models on the main tested dataset, and the repeated-run evaluations on the Baishui River and related Bazimen Z110 datasets showed relatively limited run-to-run variation under the current evaluation protocol. The MSE-TCN module improved the representation of non-fixed periodic-related patterns, while the residual-increment prediction strategy reduced the difficulty of directly modeling cumulative displacement. The interpretability analysis suggested that the model dynamically adjusted model-internal branch contributions under different deformation conditions, reflecting the varying roles of displacement history, hydrological information, and vegetation-related seasonal variation in the tested cases.

Overall, the proposed framework provides an interpretable data-driven approach for short-term reservoir-area landslide displacement prediction under sparse monitoring conditions. However, the held-out test period contains only a limited number of original monthly displacement observations, and the weekly reconstruction increases the temporal density of the sequence without increasing the number of independently observed displacement records. Accordingly, the reported performance should be interpreted within

the current retrospective reconstructed-sequence setting rather than as conclusive evidence of broad temporal generalizability or operational robustness. Broader validation on additional landslide cases, longer independently observed monitoring records, and extension to rolling-origin or near-real-time prediction with denser and more frequently updated data are still needed before stronger generalization or operational deployment claims can be made.

Author Contributions: Conceptualization, K.X. and J.Y.; methodology, J.Y. and J.S.; software, J.Y. and J.S.; validation, J.Y., J.S., K.X. and H.T.; formal analysis, J.Y. and J.S.; investigation, J.Y. and J.S.; resources, K.X.; data curation, J.Y. and J.S.; writing—Original draft preparation, J.Y. and J.S.; writing—review and editing, K.X., H.T., J.H. and W.Z.; visualization, J.Y. and J.S.; supervision, K.X.; project administration, K.X. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Natural Science Foundation of China (Grant No. 62272485 and Grant No. 62373372) and the Innovative Entrepreneurship Undergraduate Training Programme (Grant No. Yz2025471) at Yangtze University.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data used in this study were obtained from the National Cryosphere Desert Data Center.

Acknowledgments: The authors acknowledge the support of the data providers and the institutions involved in this study.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

NDVI	Normalized Difference Vegetation Index
TCN	Temporal Convolutional Network
MSE-TCN	Multi-Scale Enhanced Temporal Convolutional Network
MoE	Mixture of Experts
RMSE	Root Mean Square Error
MAE	Mean Absolute Error

Appendix A. Some Implementation Details

During training, the model was optimized using the mean squared error (MSE) between the predicted and observed standardized displacement residual increments as the primary prediction loss, together with a small MoE load-balancing regularization term ($\lambda_{lb} = 0.01$); entropy regularization was disabled.

The experiments were implemented in Python 3.10.18 using PyTorch 2.7.1, NumPy 2.0.1, pandas 2.2.3, scikit-learn 1.6.1, SciPy 1.14.1, statsmodels 0.14.5, and Matplotlib 3.9.2.

Randomness was controlled by setting the Python, NumPy, and PyTorch random seeds. The default single-run experiments, including the main comparative, ablation, visualization, and interpretability analyses, were conducted using seed 42 unless otherwise stated. The external-validation repeated-run evaluation used five random seeds: 42, 2024, 2025, 3407, and 1234. For each of the two datasets, five repeated runs were performed, resulting in ten dataset–seed combinations in total. The rolling-origin evaluation used five chronological origins, corresponding to training fractions of 60%, 65%, 70%, 75%, and 80%, under the default seed 42. The paired repeated-seed comparison between the proposed model and the

conventional TCN baseline used the same five seeds: 42, 2024, 2025, 3407, and 1234. These settings were used to distinguish the default single-run results, external-validation stability evaluation, temporal-robustness evaluation, and paired repeated-seed comparison.

Table A1. Hyperparameter and experimental settings.

Category	Item	Setting
Data processing	Data frequency	Weekly
Data processing	Train/test split	80%/20% chronological split
Data processing	Look-back window	26 time steps
Trend estimation	Method	one-sided EWMA
Optional seasonal decomposition	Nominal annual period	52 weeks; retained only for longer sequences
Prediction target	Target	Displacement residual increment
MSE-TCN block 1	Kernel size/dilation	((3, 1), (3, 2), (3, 2), (3, 4))
MSE-TCN block 2	Kernel size/dilation	((3, 1), (3, 2), (3, 2), (3, 4))
MSE-TCN	Output channels	128
MSE-TCN	Activation	ReLU
MSE-TCN	Dropout	0.2
MSE-TCN	Normalization	GroupNorm
SE module	Reduction ratio	8
Frequency modulation	STFT	(n_fft = 16), hop length = 4
Frequency modulation	Frequency bands	8
Trend branch	Encoder	One-layer LSTM
Trend branch	Hidden dimension	256
Environmental branch	Encoder	One-layer GRU
Environmental branch	Hidden dimension	256
Cross-attention	Attention dimension	128
Cross-attention	Number of heads	4
MoE	Number of experts	3
MoE	Routing	Top-k sparse routing
MoE	Top-k	2
MoE gate	Hidden dimension/activation	128/ReLU
Output head	Dimensions	512 → 384 → 256 → 128 → 1
Optimizer	Optimizer	Adam
Learning rate	Initial LR	0.001
Scheduler	Strategy	CosineAnnealingLR
Weight decay	L2 regularization	0.0005
Batch size	Batch size	32
Epochs	Maximum epochs	200
Early stopping	Patience	75
Gradient clipping	Max norm	1.0
Random seed	Default single-run experiments	42
Random seed	Paired TCN comparison and uncertainty assessment	42, 2024, 2025, 3407, 1234
Random seed	External-validation repeated-run evaluation	42, 2024, 2025, 3407, 1234
Rolling-origin evaluation	Origins and seed	60%, 65%, 70%, 75%, and 80% training fractions; seed 42

References

1. Yu, L.; Wang, Y.; Pradhan, B. Enhancing landslide susceptibility mapping incorporating landslide typology via stacking ensemble machine learning in Three Gorges Reservoir, China. *Geosci. Front.* **2024**, *15*, 101802. [\[CrossRef\]](#)
2. Liu, Y.; Xu, C.; Huang, B.; Ren, X.; Liu, C.; Hu, B.; Chen, Z. Landslide displacement prediction based on multi-source data fusion and sensitivity states. *Eng. Geol.* **2020**, *271*, 105608. [\[CrossRef\]](#)
3. Burange, R.A.; Shinde, H.K.; Mutyalwar, O. Landslide detection and mapping using deep learning across multi-source satellite data and geographic regions. *arXiv* **2025**, arXiv:2507.01123.
4. Tang, Y.; He, L.; Xiong, Q.; Xiong, Y.; Shi, Y.; Hu, G. Real-time landslide displacement prediction model based on PSO-GRU-OL-MRT. *China Saf. Sci. J.* **2025**, *35*, 70–77.

5. Ye, X.; Zhu, H.; Tian, K.; Li, H.; Zhang, W.; Cheng, G. Scientific observation and early warning of extremely large reservoir landslides from perspective of emergency management. *China Saf. Sci. J.* **2025**, *35*, 221–231.
6. Guo, Z.; Chen, L.; Gui, L.; Du, J.; Yin, K.; Do, H.M. Landslide displacement prediction based on variational mode decomposition and WA-GWO-BP model. *Landslides* **2020**, *17*, 567–583. [\[CrossRef\]](#)
7. Shao, P.; Wang, H.; Long, G.; Liao, J.; Gan, F.; Xu, B.; Hu, K.; Teng, Y. A novel multi-step ahead prediction method for landslide displacement based on autoregressive integrated moving average and intelligent algorithm. *Eng. Appl. Artif. Intell.* **2024**, *137*, 109107. [\[CrossRef\]](#)
8. Wang, L.; Yang, K.; Wu, C.; Zhou, Y.; Liu, J.; Hu, H. Novel approach to quantitative risk assessment of reservoir landslides using a hybrid CNN-LSTM model. *Landslides* **2025**, *22*, 943–956.
9. Yang, J.; Huang, Z.; Jian, W.; Robledo, L.F. Landslide displacement prediction by using Bayesian optimization–temporal convolutional networks. *Acta Geotech.* **2024**, *19*, 4947–4965. [\[CrossRef\]](#)
10. Zhang, W.; Li, H.; Tang, L.; Gu, X.; Wang, L.; Wang, L. Displacement prediction of Jiuxianping landslide using gated recurrent unit (GRU) networks. *Acta Geotech.* **2022**, *17*, 1367–1382. [\[CrossRef\]](#)
11. Nava, L.; Tordesillas, A.; Qian, G.; Catani, F. Displacement residuals reveal landslide regime shifts. *Landslides* **2025**, *22*, 1–16.
12. Wang, Y.; Tang, H.; Wen, T.; Ma, J. A hybrid intelligent approach for constructing landslide displacement prediction intervals. *Appl. Soft Comput.* **2019**, *81*, 105506. [\[CrossRef\]](#)
13. Wang, J.; Ju, Y.; Dong, P.; Ju, T. Long-term time series forecasting by a frequency-domain enhanced temporal convolutional network with the stationary residual regularization. *Appl. Soft Comput.* **2025**, *184*, 113779. [\[CrossRef\]](#)
14. Wang, S.; Wang, A.; Zhang, Y.; Xue, H.; Yao, Z. ChronoFuse-TCN: A progressive temporal convolutional network for multi-scale and spatiotemporal load disaggregation. *Energy Effic.* **2025**, *18*, 81. [\[CrossRef\]](#)
15. Fan, J.; Zhang, K.; Huang, Y.; Zhu, Y.; Chen, B. Parallel spatio-temporal attention-based temporal convolutional network for multivariate time series prediction. *Neural Comput. Appl.* **2023**, *35*, 13109–13118.
16. Xu, Z.; Zhang, Y.; Miao, Q. An attention-based multi-scale temporal convolutional network for remaining useful life prediction. *Reliab. Eng. Syst. Saf.* **2024**, *250*, 110288. [\[CrossRef\]](#)
17. Zhang, Z.; Wang, J.; Wei, D.; Xia, Y. An improved temporal convolutional network with attention mechanism for photovoltaic generation forecasting. *Eng. Appl. Artif. Intell.* **2023**, *123*, 106273. [\[CrossRef\]](#)
18. Liu, X.; Tang, Z.; Wei, J. Temporal convolutional network based on squeeze-and-excitation module and multi-scale feature fusion for GNSS positioning error correction. *GPS Solut.* **2025**, *29*, 136. [\[CrossRef\]](#)
19. Jiang, M.; Zeng, P.; Wang, K.; Liu, H.; Chen, W.; Liu, H. FECAM: Frequency enhanced channel attention mechanism for time series forecasting. *Adv. Eng. Inform.* **2023**, *58*, 102158. [\[CrossRef\]](#)
20. Huang, H.; Chen, W.; Fan, Z. TSCND: Temporal subsequence-based convolutional network with difference for time series forecasting. *Comput. Mater. Contin.* **2024**, *78*, 3665–3681. [\[CrossRef\]](#)
21. Duan, G.; Su, Y.; Fu, J. Landslide displacement prediction based on multivariate LSTM model. *Int. J. Environ. Res. Public Health* **2023**, *20*, 1167. [\[CrossRef\]](#) [\[PubMed\]](#)
22. Nava, L.; Carraro, E.; Reyes-Carmona, C.; Puliero, S.; Bhuyan, K.; Rosi, A.; Monserrat, O.; Floris, M.; Meena, S.R.; Galve, J.P.; et al. Landslide displacement forecasting using deep learning and monitoring data across selected sites. *Landslides* **2023**, *20*, 2111–2129. [\[CrossRef\]](#)
23. Wu, D.; Zhou, B.; Zimin, M. Prediction of landslide displacement based on the CA-stacked transformer model. *Alex. Eng. J.* **2025**, *124*, 389–403. [\[CrossRef\]](#)
24. Ge, Q.; Li, J.; Wang, X.; Deng, Y.; Zhang, K.; Sun, H. LiteTransNet: An interpretable approach for landslide displacement prediction using transformer model with attention mechanism. *Eng. Geol.* **2024**, *331*, 107446. [\[CrossRef\]](#)
25. Abbaszadeh Shahri, A.; Maghsoudi Moud, F. Landslide susceptibility mapping using hybridized block modular intelligence model. *Bull. Eng. Geol. Environ.* **2021**, *80*, 267–284.
26. Sun, Y.; Li, D.; Yin, K.; Chen, L.; Wang, Y. Intermittent movement prediction of colluvial landslide in the Three Gorges Reservoir: A case study of Baishuihe landslide. *Geol. Sci. Technol. Inf.* **2019**, *38*, 188–196.
27. Cruden, D.M.; Varnes, D.J. Landslide types and processes. In *Landslides: Investigation and Mitigation*; Turner, A.K., Schuster, R.L., Eds.; Transportation Research Board Special Report 247; National Academy Press: Washington, DC, USA, 1996; pp. 36–75.
28. Lu, S.; Yi, Q.; Yi, W.; Zhang, G.; He, X. Study on dynamic deformation mechanism of landslide in drawdown of reservoir water level: A case study of the Baishuihe landslide in the Three Gorges Reservoir area. *J. Eng. Geol.* **2014**, *22*, 869–875. (In Chinese) [\[CrossRef\]](#)
29. Yi, W. *Basic Characteristics and Monitoring Data of Baishui River Landslide in Zigui County, Three Gorges Reservoir Area of the Yangtze River from 2007 to 2012*; National Cryosphere Desert Data Center: Lanzhou, China, 2020. [\[CrossRef\]](#)
30. Yi, W.; Huang, H. *2007–2012 Deformation Monitoring Data of Bazimen Landslide in Zigui County, Three Gorges Reservoir Area*; National Cryosphere Desert Data Center: Lanzhou, China, 2020. [\[CrossRef\]](#)

31. Lin, Q.; Yang, Z.; Huang, J.; Deng, J.; Chen, L.; Zhang, Y. A Landslide Displacement Prediction Model Based on the ICEEMDAN Method and the TCN-BiLSTM Combined Neural Network. *Water* **2023**, *15*, 4247. [[CrossRef](#)]
32. Li, Q.; Yao, C.; Yao, X.; Zhou, Z.; Ren, K. Time Series Prediction of Reservoir Bank Slope Deformation Based on Informer and InSAR: A Case Study of Dawanzi Landslide in the Baihetan Reservoir Area, China. *Remote Sens.* **2024**, *16*, 2688. [[CrossRef](#)]
33. Chen, H.; Li, Y.; Liu, Y. Research and Analysis of the TCN-Multihead-Attention Prediction Model of Landslide Deformation in the Three Gorges Reservoir Area, China. *Front. Earth Sci.* **2025**, *13*, 1587623. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.