https://doi.org/10.5194/egusphere-2026-2352
Preprint. Discussion started: 12 May 2026
© Author(s) 2026. CC BY 4.0 License.

Reducing False Alarms in Small-Scale Slope Early Warning Systems via Deep Learning-Driven Asynchronous Displacement and Rainfall Data Fusion

Shubing Ouyang[1,2], Daichao Li[1,2], Minjiang Liu[3], Fengjian Ge[4], Xia Zheng[5], Yuan Li[6], Sheng Wu [1,2]

[1]The Academy of Digital China (Fujian), Fuzhou University, Fuzhou, 350108, China
[2]Key Laboratory of Spatial Data Mining & Information Sharing of Ministry of Education, Fuzhou University, Fuzhou, 350108, China
[3]Fujian Geologic Surveying and Mapping Institute, Fuzhou 350108, China
[4]School of Tourism and Cultural Studies and Public Administration, Fujian Normal University, Fuzhou 350108, China
[5]Fujian Province Geological Engineering Survey Center, Fuzhou 350108, China
[6]Fujian Big Data Primary Development Co., Ltd., Fuzhou 350108, China

Correspondence to: Daichao Li (lidc@fzu.edu.cn)

Abstract. Real-time crackmeter-based warning systems for rainfall-induced small-scale slope failures suffer from prohibitively high false alarm rates. Across diverse geological settings, existing approaches relying on fixed thresholds or isolated rainfall–displacement relationships are highly sensitive to noise, lacking the robustness to reliably distinguish genuine deformation signals from spurious measurements. To address this operational bottleneck, this study reframes the early warning paradigm from a traditional fixed-threshold alerting process reliant on manual discrimination to an automated alert truth verification problem, proposing a real-time binary classification framework to isolate true early warnings from sensor-triggered false alarms. Utilizing a multi-source dataset of crackmeter displacement and rainfall measurements from diverse monitored slopes in Fujian Province, China, all true early warning instances were rigorously calibrated via field investigations. A patch-based dual-branch temporally-aware Transformer model was developed to explicitly address asynchronous multi-rate data fusion, strict temporal causality between rainfall and displacement, and stringent real-time decision constraints. By simultaneously capturing long-term rainfall-displacement interactions and high-resolution displacement dynamics, the model outperforms competitive baselines, achieving a precision of 90.91%, a recall of 93.53%, and an F2-score of 92.99% in identifying true early warnings. Interpretability analysis reveals the model's decisions are primarily driven by localized displacement trends and relative rainfall intensity, aligning with expert judgment. The proposed framework significantly curtails false alarms without compromising reliability, acting as a robust decision-support layer to enhance automated slope hazard monitoring. Future work will fuse additional sensor types to suppress false alarms via cross-validation of multi-physical responses.

https://doi.org/10.5194/egusphere-2026-2352
Preprint. Discussion started: 12 May 2026
© Author(s) 2026. CC BY 4.0 License.

1 Introduction

Real-time monitoring and early warning of geological disasters form the cornerstone of disaster prevention systems, holding strategic importance for the protection of human lives and property (Khan et al., 2020; Quansah et al., 2010). Heavy rainfall frequently acts as a primary influence quantity triggering small-scale slope failures, particularly shallow landslides (Zhuang et al., 2022). These events evolve rapidly, and their precursor signals are often subtle and challenging to detect, which presents significant difficulties for the effectiveness of early warning systems. (Qiu et al., 2025). In this context, Internet of Things (IoT)-based sensing systems has become a fundamental engineering approach for continuous and automated monitoring of slope deformation. Among these sensors, crackmeters serve as core nodes within distributed IoT infrastructures, capturing real-time, high-frequency displacement data. These data represent the primary measurand associated with crack evolution, a key precursor to slope instability and structural collapse (Deng et al., 2024; Baroň and Supper, 2013; Klimeš et al., 2012).

Current methodologies for displacement-based early warning primarily rely on displacement values, velocities, or derived indicators for risk assessment. Sun et al. utilized the Mann-Kendall test to analyze evolutionary trends in fractal dimension parameters of displacement for establishing warning thresholds (Sun et al., 2024). Zhang et al. mapped normalized velocity deformation indices into a normal distribution space to achieve graded warnings (Zhang et al., 2024). However, the effectiveness of these methods depends heavily on the quality and integrity of the measurement data. In practical engineering scenarios, monitoring data from crackmeters are often contaminated by significant measurement uncertainty arising from environmental interference and inherent instrumental errors. This noise contamination directly contributes to the frequent occurrence of false positive alarms. Moreover, existing warning systems remain predominantly reliant on manual verification processes, exhibiting a lack of automated capabilities to distinguish true early warnings from false alarms. This deficiency considerably increases the demand for human analysis and judgment, imposing a substantial additional workload on monitoring personnel.

The primary challenges associated with ensuring measurement data reliability in these automated systems can be categorized as follows:

(1) Signal-to-Noise Interpretation: As a significant source of false alarms, slope displacement noise is commonly mitigated using conventional techniques like simple moving averages or Gaussian-weighted smoothing. These methods enhance data smoothness (Sharifi et al., 2024), but excessive smoothing may diminish the amplitude of critical risk signals (Sharifi et al., 2021). Furthermore, such methodologies demonstrate limited efficacy and inadequate resilience in mitigating non-stationary, transient spike noise, frequently necessitating its proactive elimination (Sharifi et al., 2021). Critically, during real-time monitoring, conventional denoising methods encounter difficulties in effectively incorporating the primary triggering factors of slope failures, such as rainfall. This makes it unreliable to determine whether a sudden change in the latest displacement value represents mere noise interference or a genuine precursor to an impending hazard.

https://doi.org/10.5194/egusphere-2026-2352
Preprint. Discussion started: 12 May 2026
© Author(s) 2026. CC BY 4.0 License.

(2) Limitations of "Prediction-Based" Validation: Given these limitations inherent in automatic noise processing, an alternative approach based on "denoising-prediction-measurement discrepancy" analysis has been proposed for true early warning recognition. This method calculates anomaly scores by quantifying deviations between denoised displacement predictions and measured values. It allows for integration with other triggering factors, and achieves superior identification of transient spike noise, thereby facilitating staged warning implementation. For instance, Shehadeh et al. employed autoregressive models to predict displacement after smoothing the displacement data, computed the statistical discrepancy between these predictions and the measured displacement, and subsequently differentiated genuine displacement events from transient spike noise (Shehadeh et al., 2024). This approach emerges within a broader evolution in landslide displacement prediction methodologies, which has progressed from early physical-mechanical models (Guo et al., 2024, 2024; Liu et al., 2024; Wu and Lan, 2019) and traditional time-series models(J. Li et al., 2021; Wang et al., 2023) to leveraging deep learning techniques (Nava et al., 2023; Wang et al., 2022; Yang et al., 2019). These advanced techniques have demonstrated an enhanced capability to extract complex nonlinear features from temporal data, significantly improving prediction accuracy. Concurrently, advances in deep learning technology have also improved the capacity of models to characterize the coupling relationships between external driving factors and displacement dynamics. This development has been demonstrated through the successful integration of multiple data sources (e.g. rainfall or reservoir water levels) into the models (Deng et al., 2021; Jiang et al., 2024), providing more substantial support for accurate displacement prediction. However, existing warning methods based on "denoise-predicted-measured discrepancy" still encounter limitations. Firstly, the process of predicting displacement typically necessitates the preliminary processing of raw data using conventional denoising methodologies (Sharifi et al., 2022). These methods often require manual parameter adjustments to accommodate the distinct characteristics of different multi-crackmeter datasets, which impedes the realization of a fully automated workflow. Secondly, displacement prediction models trained and applied on individual warning curves often demonstrate inadequate generalization capacity across multiple diverse warning curves. The prevailing approach in contemporary deep learning-based displacement prediction research involves training models on individual curves. However, this approach fails to adequately accommodate the heterogeneity of data characteristics across different geological scenarios. This shortcoming becomes particularly evident when multi-crackmeter monitoring sequences are short and lack sufficient regularity, causing models trained on single curves to underperform or fail.

In conclusion, despite the notable advancements made in the field of slope displacement noise filtering and prediction, existing techniques still face challenges in achieving automated classification of "true early warnings" and "false alarms" in multi-crackmeter real-time monitoring scenarios through a pipeline focused on a single displacement curve's "denoise-predicted-measured discrepancy". Concurrently, a significant research gap persists regarding the development of models capable of real-time, automatic differentiation of true early warnings from false alarms specifically within multi-crackmeter scenarios that are subject to multi-source noise interference. Recently, transformer-based deep temporal models have demonstrated a substantial enhancement in prediction accuracy for complex nonlinear geological displacement by meticulously focusing on critical time steps (Ge et al., 2024; Wu et al., 2025). Notably, the transformer-based PatchTST

https://doi.org/10.5194/egusphere-2026-2352
Preprint. Discussion started: 12 May 2026
© Author(s) 2026. CC BY 4.0 License.

model (Nie et al., 2022), which adopts a "patch-then-model" strategy, was utilized to mitigate noise through semantic understanding within patches (Niu et al., 2024) while simultaneously learning temporal relationships between different patches. These characteristics appear particularly well-suited to the task of true early warning recognition in multi-crackmeter scenarios. However, the functionality of PatchTST was primarily restricted to the independent extraction of features from isolated temporal channels, without explicitly capturing the cross-variable dependencies between displacement and triggering factors like rainfall. Because robust early warning recognition inherently depended on enforcing strict temporal ordering to respect the chronological precedence of meteorological precursors over slope deformation, the global attention mechanism in standard PatchTST often obscured these sequential signals. Additionally, integrating real-time, variable-frequency crackmeter data (up to 1 Hz) with hourly fixed-frequency rainfall measurements posed a significant challenge for the reliable identification of genuine failure warnings.

To address this practical challenge, this study constructs a field-verified multi-source dataset integrating crackmeter displacement and rainfall measurements from diverse monitored slopes in Fujian Province, China. Building upon this dataset, a patch-based dual-branch temporally-aware Transformer model (PatchDC) is developed to satisfy three critical engineering requirements: (1) the seamless fusion of asynchronous, multi-rate monitoring data; (2) the strict preservation of physical temporal causality between rainfall precursors and displacement responses; and (3) high-efficiency real-time applicability within operational systems. By concurrently capturing long-term rainfall-displacement interactions and high-resolution displacement dynamics, the model enables robust discrimination between genuine kinematic deformations and noise-induced false alarms.

The main contributions of this study are as follows: (1) reformulating the conventional threshold-triggered early warning process into a real-time, data-driven alarm validation task to overcome pervasive false-positive issues; (2) developing a temporally-aware dual-branch Transformer framework specifically tailored for the feature extraction and fusion of multi-rate monitoring data; and (3) demonstrating the model's operational effectiveness in significantly reducing false alarms while strictly maintaining high recall in real-world diverse monitoring scenarios.

2 The Engineering Problem: Early Warning Characteristics and System Limitations

The study area is located in Fujian Province, China, which is characterized by extensive hilly-mountainous terrain and frequent typhoon-induced rainstorms. Fujian is among China’s most geologically hazardous provinces, with disasters typically being small scale, widely distributed, highly sudden, and destructive. Typhoon-rainfall triggered landslides or collapses can progress from initial displacement to failure within minutes to hours. For example, shallow soil landslides in Wuping County (May 2022) occurred due to intense rainfall causing rapid soil saturation and reduced shear strength, leading to slope failure within minutes to hours (Li et al., 2025).

An advanced monitoring and early warning system is essential for geohazard prevention and mitigation. A 5G IoT-based geohazard monitoring and early warning system has been implemented in Fujian (Li et al., 2021), with crackmeters as a

https://doi.org/10.5194/egusphere-2026-2352
Preprint. Discussion started: 12 May 2026
© Author(s) 2026. CC BY 4.0 License.

primary monitoring device (Fig. 1). These instruments record crack displacement at variable frequencies with millimeter-level precision, enabling detection of early warning signals. True early warnings from crackmeter device alerts are confirmed only after expert review of crackmeter and rainfall data on the system platform. However, practical applications reveal a high false alarm rate: over the past five years, the platform received approximately two million device alerts, but only slightly more than 500 were validated, with just over 100 from crackmeters—an effective warning rate below 0.03%. This is mainly due to imperfect warning algorithms and anomalous data noises, such as sudden jumps and oscillations.

The crack-detection alarm algorithm currently deployed in Fujian Province relies on a simplistic, threshold-based method to enable real-time, continuous monitoring. Two primary threshold approaches are implemented: (1) The comparative analysis employs a fixed incremental threshold, where displacement values exceeding a prescribed variation limit relative to the corresponding time window of the preceding day trigger alerts. (2) Absolute displacement threshold: A fixed cumulative displacement limit is established. Exemplary threshold configurations for the four-color warning system (blue, yellow, orange, red) are detailed in Table 1.

Figure 1 On-site photos of crackmeter and rain gauge

Table 1 Exemplary threshold configurations for True early warning

Warning Level	Crack Width Change Threshold (mm)
Blue	Compared with Previous day≥20 or ≥10; The total displacement value>20
Yellow	Compared with Previous day≥40 or ≥30; The total displacement value>40
Orange	Compared with Previous day≥80 or ≥60; The total displacement value >80
Red	Compared with Previous day≥150 or ≥100; The total displacement value >150

https://doi.org/10.5194/egusphere-2026-2352
Preprint. Discussion started: 12 May 2026
© Author(s) 2026. CC BY 4.0 License.

Preliminary analysis of the confirmed true early warning data reveals that crack displacement typically exhibits two distinct characteristics: (1) a progressive upward trend over time, and (2) the discernible influence of antecedent cumulative rainfall prior to the observed displacement increase (e.g., Fig. 2). These patterns serve as critical indicators for human experts when determining warning accuracy. However, these characteristic signals are not always clearly discernible. True early warning records frequently contain anomalies within the fracture displacement data, such as abrupt noise spikes induced by equipment malfunctions (e.g., Fig. 2). A significantly more complex challenge arises from certain catastrophic events, particularly short-term, rapid landslides. These events can trigger a dramatic surge in crack displacement within minutes to an hour (e.g., Fig. 3). Critically, this abrupt change phenomenon often closely resembles data noise (e.g., Fig. 4), making differentiation based solely on displacement patterns highly problematic. In such instances, incorporating corresponding rainfall data becomes essential supplementary evidence for informed judgment. However, the currently employed threshold-based methods, which rely solely on fixed displacement thresholds, exhibit significant limitations: they often lack sensitivity to progressive displacement trends, fail to incorporate antecedent rainfall conditions, and struggle to reliably distinguish genuine deformation signals from noise, resulting in high false positive rates. By contrast, their key advantage lies in achieving a high recall rate, ensuring minimal missed detections of critical hazards. This inherent capability to capture essential events despite pervasive noise motivates our exploration of deep learning techniques that build upon the high-recall characteristic of threshold-based approaches. The ultimate goal is to substantially reduce false alarms while maintaining a comparable recall performance.

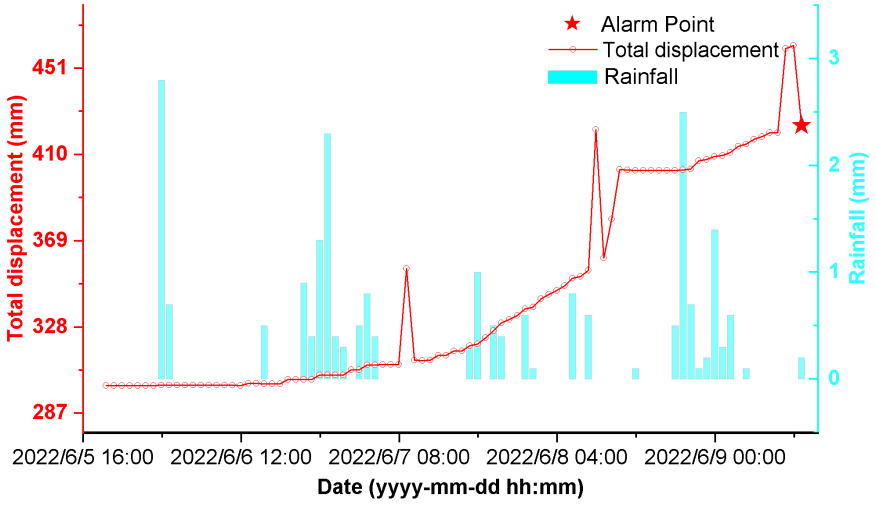


Figure 2 True early Warning: Progressive fracture displacement patterns and cumulative rainfall impact mechanisms

https://doi.org/10.5194/egusphere-2026-2352
Preprint. Discussion started: 12 May 2026
© Author(s) 2026. CC BY 4.0 License.

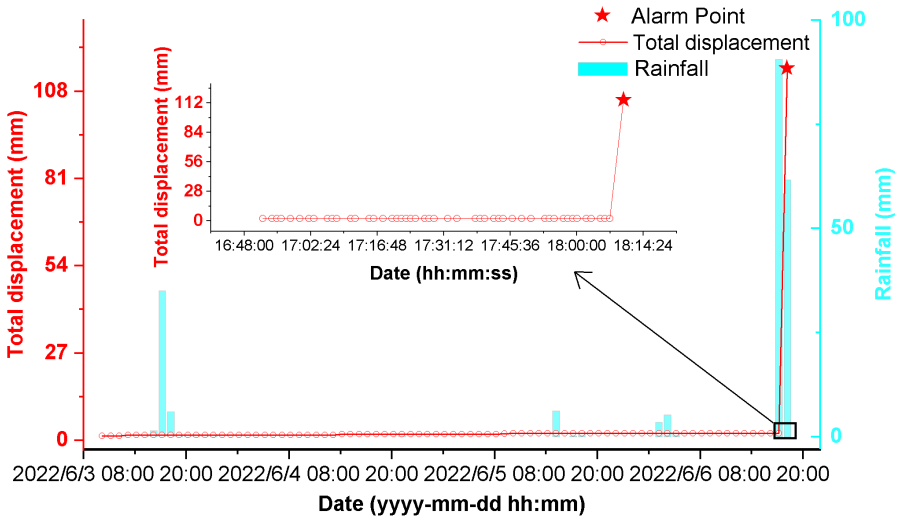


Figure 3 True early Warning: Short-term landslide fracture displacement surge induced by torrential rainfall (dramatic change within 3 minutes)

170

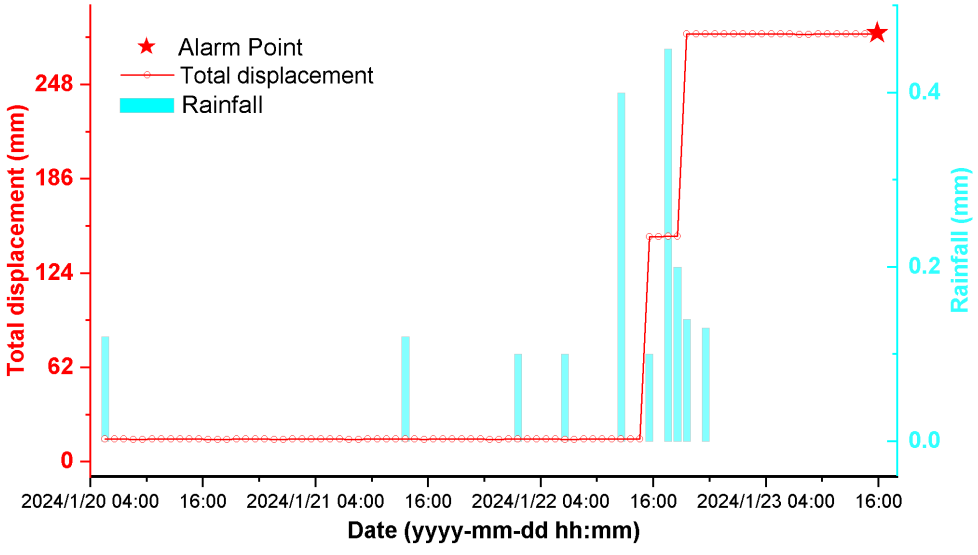


Figure 4 False alarm: Abnormal fracture displacement spike and persistence under low cumulative rainfall (merely 2.24mm of rainfall accumulated in the 10 days prior to the warning)

https://doi.org/10.5194/egusphere-2026-2352
Preprint. Discussion started: 12 May 2026
© Author(s) 2026. CC BY 4.0 License.

3 Towards an AI-Driven Solution: Problem Definition and Dataset Construction

3.1 Problem Definition and Methodological Framework

The core objective of this study was to utilize historical crack displacement and synchronous rainfall data from integrated crack-rain gauges to assess, in real time, whether an alarm generated by the latest displacement measurement signified a genuine event or a false trigger. Here, the generated alarm explicitly referred to equipment warnings issued automatically by crackmeters using the fixed-threshold method described earlier. Specifically, the task was formally framed as a binary classification problem: determining whether the latest data point (x_1) in the displacement time series (which triggered the alarm) signified a true early warning or a false alarm. A true early warning was strictly defined as one that had been subsequently field-verified by geological experts as indicating a genuine geohazard risk. The primary objective of this classification was to significantly reduce the false alarm rate of the monitoring equipment while maintaining a very high recall rate, thus ensuring safety was not compromised. To provide a clear roadmap of this study, Fig. 5 depicts the

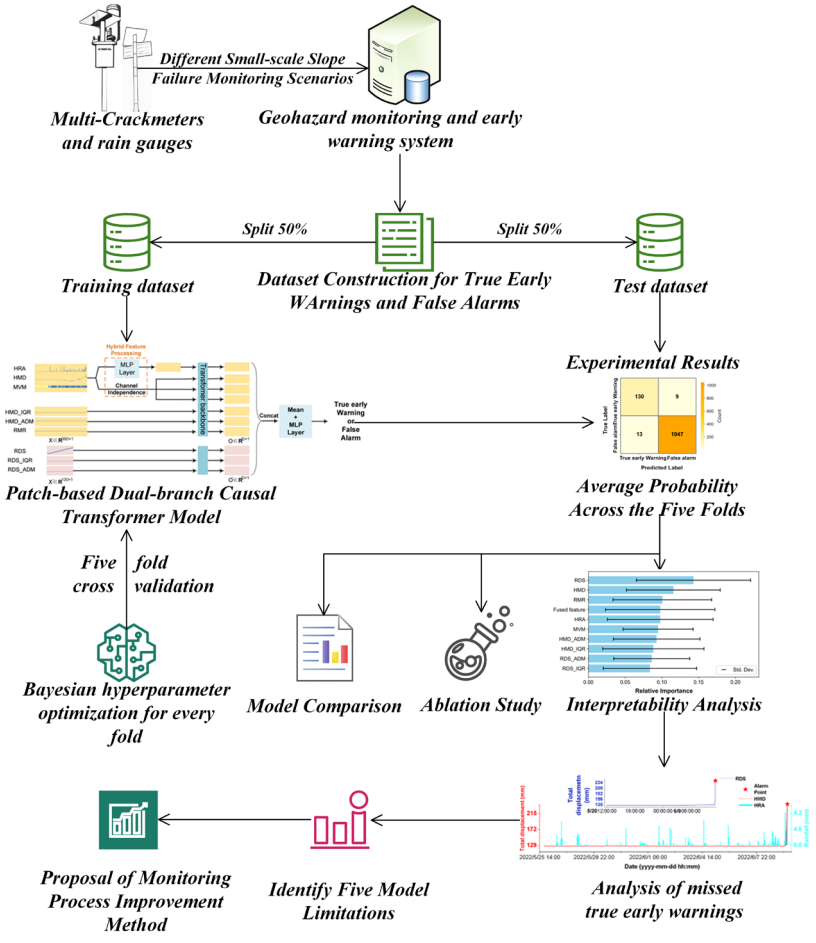


Figure 5 Workflow Overview

https://doi.org/10.5194/egusphere-2026-2352
Preprint. Discussion started: 12 May 2026
© Author(s) 2026. CC BY 4.0 License.

comprehensive workflow, which is structured into four cohesive phases: (1) Data Preparation and Preprocessing, (2) Model Training and Evaluation, (3) Model Comparison and Ablation Study, and (4) Interpretability Analysis, Limitations Analysis and Workflow Enhancement.

3.2 Dataset Construction

All data was obtained from 345 landslide or collapse masses, with 342 being small-scale slopes. Field investigations confirmed rainfall as the dominant trigger, with 105 identified masses exhibiting thicknesses below 2 m, accounting for 30% of the total. The data collection and processing procedures consisted of two categories: (1) displacement data from crackmeters, and (2) rainfall data from rain-intensity gauges.

(1) Displacement Data. Displacement data was collected from 429 crackmeters deployed across Fujian Province, covering the period from 2021 to 2024 (Fig. 6). These crackmeters featured a high maximum accuracy of 0.1 mm. Data acquisition employed a variable frequency sampling strategy, with the maximum sampling rate reaching 1 Hz. The sampling frequency was typically elevated automatically during intense rainfall events or upon detection of significant displacement changes. From this continuous monitoring dataset, 379 sensors yielded a large pool of 2142 false alarm instances. Conversely, 93 true early warning instances originating from the remaining 52 sensors were strictly confirmed through rigorous on-site field verification. Because this initial quantity was insufficient for robust deep learning model training, and given that device-generated warnings were triggered deterministically by fixed thresholds (e.g., a displacement change from 50 to 55 mm generating a new warning), data points temporally adjacent to these confirmed events were treated as potential true warnings. Subsequently, randomly selected points in the vicinity of these verified warnings underwent a rigorous manual review. Specifically, 162 adjacent instances were subjected to a triple-blind review by three independent geological experts. Only the points that received unanimous agreement from all three experts were incorporated. This procedure successfully increased the total number of field-validated true early warning instances to 255, demonstrating high label reliability. Furthermore, to address potential concerns regarding temporal leakage among these highly correlated adjacent samples, all data originating from the same sensor or specific geological event were strictly grouped into either the training or the testing set. This sensor-level isolation prevented highly correlated adjacent segments from cross-contaminating the evaluation phase, fundamentally neutralizing the risk of data leakage.

The raw displacement data from crackmeters was processed into the following seven distinct features (Table 2):

Hourly Median Displacement. Based on established research indicating that landslides are most strongly associated with precipitation effects within a 6-10 day window (Dahal and Hasegawa, 2008; Dou et al., 2024; Lee et al., 2021), a 360-hour (15-day) retrospective window preceding the warning timestamp was selected. To reduce high-frequency noise in the displacement data and ensure temporal alignment with hourly rainfall measurements, the median displacement was calculated for each hourly interval within this window, excluding the final hour. For the terminal hour (the hour containing the warning), the raw sensor-generated warning value was retained instead of the median to preserve the temporal immediacy and raw signal characteristics essential for accurate warning validation. If no valid displacement data existed

https://doi.org/10.5194/egusphere-2026-2352
Preprint. Discussion started: 12 May 2026
© Author(s) 2026. CC BY 4.0 License.

within a specific hour, the gap was filled via interpolation using the median value from the most recent preceding valid hourly interval. Furthermore, it was recognized that not all sensors provided continuous data coverage for the full 360-hour window, missing early time segments were pragmatically backfilled with the last available valid measurement from the respective sensor until the sequence length reached 360 hours. Finally, all resulting 360-hour displacement sequences underwent mean normalization prior to being input into the model. To avoid artificial stationarity or signal suppression introduced by backfilling, this work combined explicit uncertainty masking with architectural decoupling. The Missing Value Mask explicitly flagged imputed segments, preventing misinterpretation of flat backfilled intervals as slope stability. Concurrently, the High-Resolution branch preserved raw, non-interpolated displacement data intact. Last-valid-measurement backfilling served only for temporal alignment and never contaminated the high-resolution signal path. Detailed descriptions of these features were provided in the subsequent sections.

Missing Value Mask. A complementary binary mask was generated corresponding directly to the Hourly Median Displacement sequence. This mask indicated data presence (1) or absence (0) for each hourly bin, thereby informing the model where imputed values occurred and potentially aiding its interpretation.

Raw Displacement Sequence. Considering that the progressive deformation characteristics are critical for early warning within this sub-hourly timeframe. To preserve this detailed displacement information immediately preceding sensor-triggered alerts, a sequence of 100 consecutive high-resolution raw measurements immediately prior to (and including) the warning displacement value was adopted. These data volume ensured comprehensive coverage of at least the final hour before warning activation, capturing potential rapid changes. All raw data in this sequence underwent mean normalization prior to being used.

The interquartile range and the absolute displacement magnitude (four statistical features). To reintroduce valuable physical scale information that was obscured by the prior normalization steps, four statistical features were derived from each input sequence. For the hourly median displacement sequence, these included: 1) the interquartile range (IQR), which mitigated sensitivity to extreme noise values, and 2) the absolute displacement magnitude, calculated as the alert value minus the minimum value within the sequence. Analogous features—IQR and absolute displacement magnitude—were similarly extracted from the high-resolution raw displacement sequence. All four derived statistical features were subsequently normalized by their collective maximum value across the dataset. To ensure compatibility with the model input format, each normalized statistical feature was replicated to match the length of its corresponding input sequence (360 values for features related to hourly medians; 100 values for raw displacements).

(2) Rainfall Data. Regarding rainfall data, measurements were predominantly obtained from rain gauges installed in proximity to crackmeters. These rain gauges have a measurement resolution of 0.1 mm. The predominant sampling frequency is hourly, although a small proportion of datasets record at sub-hourly intervals. Rainfall data was processed into the following two features:

Single-hour rainfall accumulation. Rainfall values were summed or extracted according to the time period represented by each Hourly Median Displacement value. Before use, the data was subjected to min-max normalization.

https://doi.org/10.5194/egusphere-2026-2352
Preprint. Discussion started: 12 May 2026
© Author(s) 2026. CC BY 4.0 License.

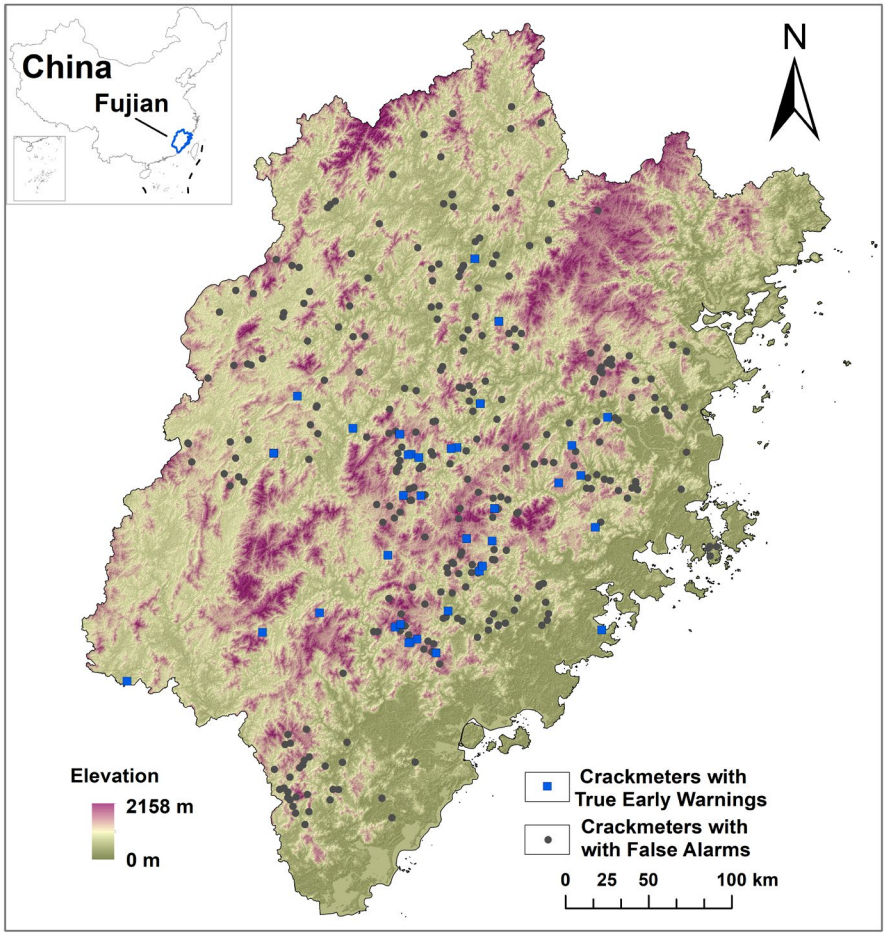


Figure 6 Crackmeters Spatial Distribution Map

Relative Maximum Rainfall Intensity. After normalization, the original rainfall data lost its absolute physical meaning, thereby making direct comparisons of rainfall amounts between different regions difficult. Furthermore, evidence suggests that the risk of geological disasters may be significantly higher if the rainfall intensity is substantially increased relative to their own historical levels (Li et al., 2025).

Therefore, to characterize the maximum rainfall intensity for each region within the observational temporal window, both relative to other regions (spatially) and relative to its own history (temporally), the following key steps were implemented:

A. For each rain gauge time series comprising 360 data points, the maximum value was extracted (Single-hour rainfall_max_i).

B. The global maximum value was extracted from all individual Single-hour rainfall_max_i (Single-hour rainfall_max_all).

C. The average annual rainfall (Annual rainfall_i) for the preceding five years for the region corresponding to the individual rain gauge was extracted. This annual rainfall data were calculated based on the high-resolution 1-km monthly precipitation

https://doi.org/10.5194/egusphere-2026-2352
Preprint. Discussion started: 12 May 2026
© Author(s) 2026. CC BY 4.0 License.

dataset for China (1901–2024), which is provided by National Tibetan Plateau Data Center (https://data.tpdc.ac.cn/zh-hans/data/faae7605-a0f2-4d18-b28f-5cee413766a2/) (Peng et al., 2019).

D. The maximum average annual rainfall (Annual rainfall_max) observed across all regions over the last five years was extracted.

E. Dual normalization was performed using the above values. The core formula is defined as follows:

$$\mathrm{rain}^{\mathrm{i}}_{stastic} = \frac{\text{Single} - \text{hour rainfall\_max\_i} \times \text{Annual rainfall\_max}}{\text{Single} - \text{hour rainfall\_max\_all} \times \text{Annual rainfall\_i}} \quad (1)$$

Following the aforementioned calculation, Relative Maximum Rainfall Intensity was defined as a dimensionless index to normalize precipitation extremity across heterogeneous climatic zones. By scaling local peak intensities against regional climatological baselines, this metric accounted for the heightened sensitivity of relatively arid regions to anomalous rainfall events. This transformation facilitated robust cross-regional pattern recognition by characterizing the relative extremity of events rather than absolute physical volume, ensuring consistent feature distributions across diverse geographical scenarios. To ensure compatibility with the model input format, the derived statistical value was expanded into a 360-point sequence by replicating its value, thereby aligning it with the temporal resolution of the single-hour rainfall accumulation data.

4 Method

4.1 Measurement Data Processing Framework

The proposed model employed a dual-branch architecture, leveraged a modified PatchTST structure for multi-rate sampling challenge (PatchDC, Fig. 7). The first branch, termed the Long-Term Rainfall-Displacement Modeling Branch, was designed to capture displacement trends and the impact of rainfall on displacement magnitude at relatively long 15-day scales with an hourly resolution. The second branch, the High-Resolution Crack Displacement Branch, was tasked with capturing very fine-grained minute to second-resolution displacement trends directly from raw sensor data. The fused features extracted from both branches were subsequently fed into a classification module to determine the warning category. All hyperparameters of the model were optimized via Bayesian inference. The specific hyperparameters and their optimized values are detailed in Table 3.

(1) Long-Term Rainfall-Displacement Modeling Branch

Its input comprised six features: HMD, MVM, HRA, HMD_IQR, HMD_ADM, RMR (Table 2). The module architecture primarily consisted of a Hybrid Feature Processing followed by a Transformer Backbone, with technical specifications that were detailed below.

https://doi.org/10.5194/egusphere-2026-2352
Preprint. Discussion started: 12 May 2026
© Author(s) 2026. CC BY 4.0 License.

Table 2 The all features of the datasets

Source	Feature(abbreviation)	Seq. Len.	Resolution	Description
Crackm-eter (Measurand)	Hourly Median Displacement (HMD)	360	Hourly	Median displacement per hour in 360-hr window preceding warning (terminal hour: raw warning value).
	Missing Value Mask (MVM)	360	Hourly	Binary indicator (1=valid, 0=missing) for hourly displacement bins.
	Raw Displacement Sequence (RDS)	100	⩽1 Hz	100 Raw Displacement prior to/including warning value.
	Interquartile range for HMD (HMD_IQR)	360	Hourly	Interquartile range of 360-hr sequence (noise robustness).
	Absolute Displacement Magnitude for HMD (HMD_ADM)	360	Hourly	Alert value minus min value in 360-hr sequence.
	Interquartile range for RDS (RDS_IQR)	100	≤1 Hz	Interquartile range of 100-point raw sequence.
	Absolute Displacement Magnitude for RDS (RDS_ADM)	100	≤1 Hz	Alert value minus min value in raw sequence.
Rain Gauge (Influence Quantity)	Single-hour rainfall accumulation (HRA)	360	Hourly	Hourly rainfall aligned with displacement window
	Relative Max Rainfall Intensity (RMR)	360	Hourly	The rainfall intensity relative to both its spatial context and its own history.

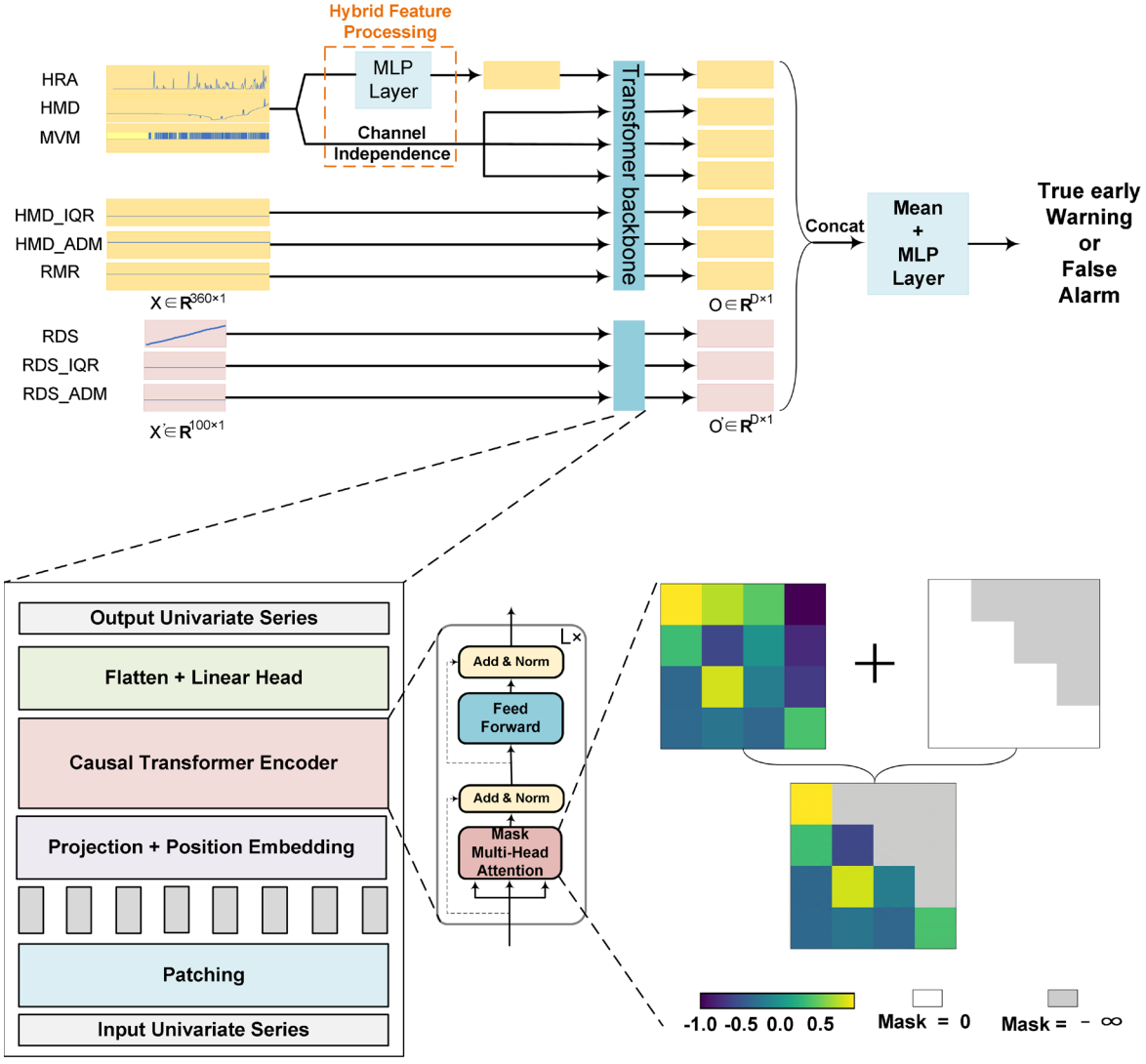
Table 3 The hyperparameter search space of PatchDC

Parameters	Hyperparameter search space
hidden_size	32, 64, 128
encoder_layers	1, 2, 3
n_heads	2, 4, 8
patch_len	8, 12, 16, 20, 24
dropout	0.1, 0.3, 0.5
stride	3, 5, 7

Hybrid Feature Processing. Unlike the original PatchTST model's strict channel-independent processing approach, which treated each time series variable separately while ignoring explicit correlations between variables, this study adopted a simplified correlation strategy specifically between rainfall and displacement. Specifically, three correlated bands ("HRA", "HMD", "MVM") were jointly processed through a shared multilayer perceptron layer (MLP): an initial linear transformation along the channel dimension was followed by linear rectification function and linear projection, which generated a fused feature representation that captured rainfall-displacement interactions. This integrated feature representation was subsequently combined with all remaining long-term features, thereby preserving the model's capacity for univariate pattern recognition ($X \in R^{360\times1}$) while significantly enhancing its ability to model cross-variable temporal correlations.

Transformer Backbone (Fig. 7). Based on these components, this study further incorporated them into the transformer backbone of the PatchTST method by replaced the bidirectional self-attention with masked self-attention (Vaswani et al.,

https://doi.org/10.5194/egusphere-2026-2352
Preprint. Discussion started: 12 May 2026
© Author(s) 2026. CC BY 4.0 License.



315 **Figure 7 Overall structure of the proposed PatchDC model.**

2017). A crucial modification ensured temporal ordering by architecturally constraining patch attention to historical contexts only, thereby physically respecting the precedence of precursor events (e.g., rainfall) over target phenomena (e.g., landslide displacement) in classification tasks. The transformer backbone of PatchTST and its specific architecture are illustrated in
320 Fig. 7. The core process involved segmenting each univariate time series into patches, followed by temporal feature learning through transformer modules, with a detailed description described below.

A. Patching and Projection. Given a univariate time series $X \in \mathbb{R}^{360\times1}$ taken from one component, a sliding window partitioned X into N overlapping patches:

https://doi.org/10.5194/egusphere-2026-2352
Preprint. Discussion started: 12 May 2026
© Author(s) 2026. CC BY 4.0 License.

$$\mathbf{X}_p^{(i)} = \mathbf{X}_{(i\cdot S):(i\cdot S+P)} \quad \forall i \in \{0,1,\ldots,N-1\} \tag{2}$$

where P (patch_len, Table 3) denoted patch length, S (stride, Table 3) was the sliding stride, and N = [(L-P)/S] + 2. The patches $\mathbf{X}_p \in \mathbb{R}^{\mathrm{P}\times\mathrm{N}}$ was linearly projected into a latent space $\mathbf{E}_p \in \mathbb{R}^{\mathrm{D}\times\mathrm{N}}$, where D (hidden_size, Table 3) was the Bayesian-optimized hidden dimension.

B. **Positional Embedding.** Learnable positional embeddings $\mathbf{P}_p \in \mathrm{R}^{\mathrm{D}\times\mathrm{N}}$ were element-wise added to the latent space E_p to form representation Z_p, enabling the model to explicitly capture the temporal ordering of rainfall and displacement patches.

C. **Temporally-aware Transformer Encoder.** The encoder stack consisted of L identical layers (encoder_layers, Table 3), each that implemented a masked multi-head attention mechanism. For each attention head h=1,...,H (n_heads, Table 3), the input sequence Z_p underwent three linear transformations to generate:

$$\mathbf{Q}_h = (\mathbf{Z}_P)^T\,\mathbf{W}_h^Q, \quad \mathbf{K}_h = (\mathbf{Z}_P)^T\,\mathbf{W}_h^K, \quad \mathbf{V}_h = (\mathbf{Z}_P)^T\,\mathbf{W}_h^V \tag{3}$$

$$\mathbf{W}_h^Q, \mathbf{W}_h^K \in \mathbb{R}^{D\times D_k},\ \mathbf{W}_h^V \in \mathbb{R}^{D\times\mathrm{D}},\ D_k = \mathrm{D}/\mathrm{H} \tag{4}$$

Subsequently, a lower-triangular mask matrix M. M was applied to the projected $\mathbf{Q}_h$, $\mathbf{K}_h$, and $\mathbf{V}_h$ tensors to enforce temporal ordering. The M and the masked self-attention output for each head (head_h) were computed as:

$$\mathbf{M}_{ij} = \begin{cases} 0 & j \le i\ (historical\ contexts) \\ -\infty & j > i\ (future\ blocks) \end{cases} \tag{5}$$

$$\mathrm{head}_h = \mathrm{softmax}\left(\frac{\mathbf{Q}_h\mathbf{K}_h^{\mathrm{T}}}{\sqrt{D_k}} + \mathbf{M}\right)\mathbf{V}_h \tag{6}$$

The outputs of all attention heads were subsequently processed through a feature transformation block, which consisted of layer normalization, residual connections, and a position-wise feedforward network (FFN). The FFN was composed of two linear layers with a Gaussian Error Linear Unit (GELU) activation function in between. The final output of this block constituted the embeddings of the Temporally-aware Transformer Encoder.

D. **Flatten and Linear Head.** The encoder embedding results were flattened and projected via a linear head to produce output of univariate series $\mathrm{O} \in \mathbb{R}^{\mathrm{D}\times 1}$.

(2) High-Resolution Crack Displacement Branch

This high-resolution branch was specifically designed to overcome the primary branch's limitation in detecting sub-hourly variations by processing raw crack displacement signals at their native temporal granularity. Its input comprised three features: RDS, RDS_IQR, RDS_ADM ($\mathrm{X}' \in \mathbb{R}^{100\times 1}$). It employed the identical Transformer backbone architecture with masked self-attention as described for the primary branch.

(3) Multiscale Feature Fusion Framework

https://doi.org/10.5194/egusphere-2026-2352
Preprint. Discussion started: 12 May 2026
© Author(s) 2026. CC BY 4.0 License.

The outputs from both branches were concatenated, followed by dropout regularization (dropout, Table 3) to mitigate overfitting. A mean operation was then performed along the final feature dimension for dimensionality reduction. This was succeeded by a fully connected layer that reduced the features to two dimensions, softmax activation, and was integrated into the cross-entropy loss function.

4.2 Training Strategy, Environment and Parameters

Model training employed a five-fold cross-validation strategy. To ensure optimal performance for each fold, Bayesian hyperparameter optimization was independently conducted for every fold using the training data. The optimization utilized the Optuna 3 framework with the Tree-structured Parzen Estimator (TPE) as the sampler. Thirty optimization trials were executed on the training set of each fold, with the validation set F_2-score serving as the optimization objective to align with the goal of enhancing recall for true early warnings. The hyperparameter search space for key parameters is detailed in Table 3. Additionally, the learning rate was uniformly sampled in logarithmic space within the range of 1×10^{-3} to 1×10^{-2}. The batch size was fixed at 100. To address the class imbalance, a weighted cross-entropy loss function was employed. The positive-to-negative sample weight ratio was set to 10:1, strictly aligning with the actual class distribution ratio observed in the training dataset. This data-driven weighting strategy ensured that the model did not become biased toward the majority class (false alarms), thereby preserving its sensitivity to the rare but critical true early warnings and minimizing the risk of missed detections. Models were trained using the cross-entropy loss function for a maximum of 200 epochs per fold. An early stopping strategy was implemented which monitored the validation set F_2-score; training was halted if no improvement greater than Δ=0.001 was observed for 40 consecutive epochs (patience=40). All training was performed on an NVIDIA GeForce RTX 3090 GPU (24 GB VRAM) running the Ubuntu 24.04.1 LTS operating system. During the final test phase, the posterior probabilities for the "true early warning" class output by the five independent models (from the five folds) were arithmetically averaged. Test samples were then classified as "true early warning" if this average probability was ≥ 0.5; otherwise, they were classified as "false alarm". The entire process of training and testing on the five folds was run five separate times with different random seeds to validate the reproducibility and stability of the model's accuracy.

The dataset was split into training-validation and test sets via random sampling, with an approximate ratio of 1:1. Samples derived from true real warnings were partitioned strictly by sensor ID (not by alarm event). This meant all alerts from a given sensor were assigned entirely to either the training or the test set, ensuring no data leakage. Sensors were randomly assigned to sets. Detailed splitting results, including counts per set, are provided in Table 4.

Table 4 Division of Training, Validation, and Test Sets

	Alert events of True warnings	Alert events of False warnings
Training-validation set	116	1082
Testing set	139	1060

https://doi.org/10.5194/egusphere-2026-2352
Preprint. Discussion started: 12 May 2026
© Author(s) 2026. CC BY 4.0 License.

4.3 Evaluation Criteria for Measurement Reliability

In the context of geological safety monitoring, standard classification accuracy is insufficient. We adopt specific metrics to quantify the reliability and trustworthiness of the measurement system:

(1) Recall (Recall for the true warning class): This is the most critical metric for safety, measuring the system's ability to correctly identify all genuine slope deformation events (avoiding missed alarms).

(2) Precision (Signal Validity): This metric reflects the False Alarm Rate (FAR). High precision implies that when the system issues a warning, it is highly likely to be physically significant, thereby reducing operator fatigue from nuisance alarms.

(3) F_2-score: A composite metric that weighs Recall higher than Precision, aligning with the "safety-first" principle of geological instrumentation.

(4) AUPR (Area Under the Precision-Recall Curve): Used to evaluate the stability of the measurement validation across different decision thresholds, particularly suitable for imbalanced datasets where true warnings are rare events.

5 Experimental Results

The median recall result for the true early warning class was selected from the five repeated experiments as the representative performance measure. The testing results indicated an overall accuracy of 98.17%. For the true early warning class specifically, precision reached 90.91%, recall was 93.53%, and the F_2-score was 92.99%. As shown in the confusion matrix (Fig. 8), there were 9 instances of missed detections of true early warnings (False Negatives) and 13 instances where false alarms were incorrectly (False Positives). Detailed case analyses of these missed detections and potential mitigation methods are discussed in Section 6.4.

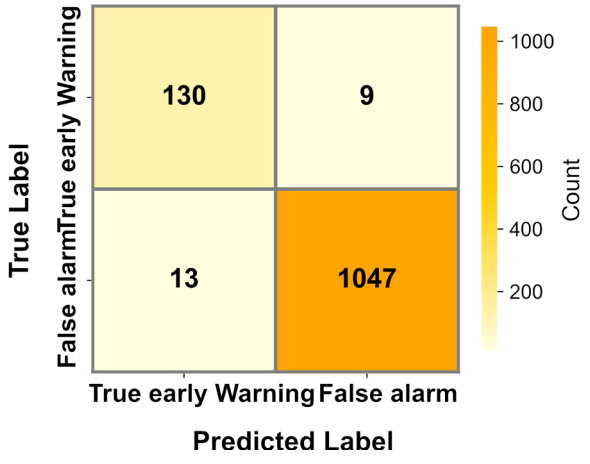


Figure 8 Confusion Matrix for PatchDC Model Test Results

https://doi.org/10.5194/egusphere-2026-2352
Preprint. Discussion started: 12 May 2026
© Author(s) 2026. CC BY 4.0 License.

6 Discussion

6.1 Comparative Model Accuracy Analysis

To illustrate the performance of PatchDC, this study compared it against multiple models applicable to time series classification. These include: the Bayesian Intensity-Duration (I-D) Threshold, which serves as a traditional empirical baseline by establishing the minimum meteorological conditions (rainfall intensity and duration) necessary to trigger slope failures using objective statistical inference (Guzzetti et al., 2007); Long Short-Term Memory (LSTM), which excels at capturing long-term dependencies in sequences and remains one of the most widely adopted models for landslide displacement prediction (Hochreiter and Schmidhuber, 1997; Taorui et al., 2022; Xu and Niu, 2018); Temporal Convolutional Network (TCN), which emphasizes temporal ordering through causal convolutions for efficient long-sequence processing and has gained significant traction in landslide displacement prediction in recent years (Bai et al., 2018; Luo et al., 2022; Yang et al., 2024); TimesNet, a newer model that transforms time series into the frequency domain for multi-periodicity modeling (Wu et al., 2022); and DLinear, a simple yet efficient architecture that performs forecasting through decomposition into trend and seasonal components (Zeng et al., 2023). Furthermore, this study contrasted recent Transformer-based enhanced models, including PatchTST and iTransformer. PatchTST processes time series by segmenting them into sub-series patches, improving local feature capture while reducing computational complexity (Nie et al., 2022); iTransformer inverts the standard attention mechanism to focus on variable dimensions rather than time steps, which is reported to enhance the accuracy of multivariate time series forecasting (Liu et al., 2023). All comparative models were adapted to employ a dual-branch input structure analogous to PatchDC to ensure a fair comparison. One branch processed the long-term features (HMD, MVM, HRA, HMD_IQR, HMD_ADM, RMR), while the other branch processed the high-resolution features (RDS, RDS_IQR, RDS_ADM). The features from the two branches were simply concatenated, projected through a linear layer, and then forwarded to the loss function. To ensure fair training, Bayesian hyperparameter optimization was employed for each fold and the specific hyperparameters searched for each are detailed in Table 5. The specific training environment and optimization method are described in Section 4.2. Each model was trained and evaluated five times, with the mean and standard deviation of the metrics reported as the final results (Table 5).

As shown in the Table 5, the optimal model, PatchDC, outperformed the suboptimal PatchTST model by margins of 3.62%, 0.43%, 1.13%, and 0.55% in Precision of true early warning, Recall of true early warning, F_2-score of true early warning, and Overall accuracy, respectively. This demonstrated that the proposed Hybrid Feature Processing mechanism and the introduced masked self-attention module effectively enhanced proposed model's capability. TimesNet ranked third, underperforming relative to PatchDC with lower scores in Precision of true early warning by 5.79%, Recall by 1.15%, F_2-score by 4.9%, and Overall accuracy by 0.95%. This performance gap may be attributed to its extensive feature decomposition into multiple periods, which might inadvertently discard critical discriminative features that are salient for

https://doi.org/10.5194/egusphere-2026-2352
Preprint. Discussion started: 12 May 2026
© Author(s) 2026. CC BY 4.0 License.

short-term warning recognition. The iTransformer model lagged notably behind PatchDC across all accuracy metrics, with margins ranging from 2.73% to 14.92%. Both iTransformer and PatchDC employed a Transformer architecture; however, their core focus differed significantly: iTransformer focused on modeling intra-variable relationships across the feature dimension, while PatchDC prioritized capturing intra-variable temporal patterns along the time dimension, which appeared more suited for this task. In contrast, the DLinear model excelled at capturing univariate trend features, resulting in a real alarm Recall rate nearly comparable to PatchDC (merely 0.29% lower). However, DLinear lacked the capability analogous to PatchDC to leverage the Transformer architecture for interpreting complex semantic patterns within local temporal segments. This limitation potentially resulted in lower robustness to noise and explained why DLinear significantly underperformed PatchDC in false alarm detection (Precision), with a performance gap of 23%. Models commonly used for landslide displacement prediction, LSTM and TCN, primarily focused on learning multivariate feature correlations but tended to neglect the importance of detailed univariate temporal characteristics for prediction accuracy. Thus, LSTM underperformed PatchDC by 13.99%, 4.75%, 6.95%, and 2.63% in Precision, Recall, F_2-score, and Overall accuracy, respectively. Similarly, TCN lagged by 14.29%, 3.16%, 5.82%, and 2.55% in these same metrics. Furthermore, to establish a traditional empirical baseline, the Bayesian Intensity-Duration (I-D) Threshold model, which relies exclusively on rainfall data, was also evaluated. While this model achieved the highest Recall (97.12%) among all evaluated methods—reflecting its highly conservative, safety-first nature—it exhibited severe deficits in Precision (20.87%), F_2-score (56.11%), Overall accuracy (57.11%), and AUPR (20.60%). This exceedingly high false alarm rate underscores the fundamental limitation of traditional meteorological thresholds: by ignoring the actual kinematic response of the slope (displacement) and relying solely on triggering factors, they lack the capacity to reliably verify genuine deformation signals. This stark contrast further highlights the operational superiority and necessity of PatchDC's multi-source data fusion framework.

The computational complexity of the algorithm was not explicitly analyzed in this study. This is because the displacement data from crack meters and the rainfall data from rain gauges are transmitted to the Fujian Provincial Integrated Geological Disaster Management Platform, where the algorithm can be deployed on computationally powerful infrastructure. Although the Transformer-based model entails higher computational cost and memory usage, the entire process from preliminary data processing to prediction using the trained model requires only a few seconds. This is remarkable despite the fact that the computational resources used for training were more limited than those on the deployment platform. This indicates that PatchDC's performance on the platform is fully adequate to meet the requirements for real-time prediction.

6.2 Ablation Study

To validate the efficacy of the two proposed core modules (Hybrid Feature Processing and Masked Self-Attention), ablation studies were conducted. The baseline was the standard PatchTST model adapted to our dual-branch input but without these modules. As shown in Table 6, integrating only the Hybrid Feature Processing mechanism into this adapted PatchTST

https://doi.org/10.5194/egusphere-2026-2352
Preprint. Discussion started: 12 May 2026
© Author(s) 2026. CC BY 4.0 License.

Table 5 Accuracy results of different models and the Hyperparameter search space of models

Model	Input data/structure	Precision of true early warning	Recall of true early warning	F_2-score of true early warning	Overall accuracy	AUPR
Bayesian I-D Threshold	Rainfall data	20.87	97.12	56.11	57.11	20.60
TCN	Dual-branch	74.95±1.64	90.22±0.35	86.68±0.44	95.36±0.30	88.78±0.84
LSTM	Dual-branch	75.25±3.00	88.63±1.95	85.55±1.56	95.28±0.54	89.08±1.60
TimesNet	Dual-branch	83.45±2.86	92.23±0.29	87.60±1.55	96.96±0.43	95.68±0.66
DLinear	Dual-branch	66.24±3.04	93.09±2.64	86.02±1.06	93.64±0.64	82.75±1.94
PatchTST	Dual-branch	85.62±1.80	92.95±1.24	91.37±0.63	97.36±0.16	95.58±0.22
iTransformer	Dual-branch	74.32±2.97	89.50±3.33	85.95±2.71	95.18±0.65	87.77±1.33
PatchDC	Dual-branch	89.24±2.43	93.38±0.95	92.50±0.49	97.91±0.28	96.35±0.54
The Hyperparameter search space	Bayesian I-D Threshold: Inter-event Time Definition (hours)：[6,12,24,32]，The prior distributions for parameters alpha and beta, as well as the formulation of the power-law threshold I = alpha D^{-beta}, were configured following the frequentist-Bayesian framework established by (Guzzetti, 2007).					
	TCN: num_levels: [8,16,32,64]; dropout: [0.1, 0.3, 0.5]; branch_channels: [1 × num_levels, 2 × num_levels, 4 × num_levels];					
	LSTM: hidden_size: [32,64, 128]; num_layers: [1, 2, 3]; dropout: [0.1, 0.3, 0.5];					
	TimesNet: top_k: [2, 3, 4, 5]; d_ff: [8,16, 32]; num_kernels: [3,5,7]; d_model: [32,64, 128]; e_layers: [1,2, 3]; dropout: [0.1, 0.3, 0.5];					
	DLinear: moving_avg_window: [15, 21, 27, 33, 39, 45]; pred_len: [1, 12, 24, 48]; dropout: [0.1, 0.3, 0.5];					
	PatchTST: hidden_size: [32, 64, 128]; encoder_layers: [1, 2, 3]; n_heads: [2, 4, 8]; patch_len: [8, 12, 16, 20, 24]; dropout: [0.1, 0.3, 0.5]; stride: [3, 5, 7];					
	iTransformer: d_model: [32,64, 128]; e_layers: [1,2, 3]; n_heads: [2, 4, 8]; dropout: [0.1, 0.3, 0.5];					
	PatchDC: hidden_size: [32, 64, 128]; encoder_layers: [1, 2, 3]; n_heads: [2, 4, 8]; patch_len: [8, 12, 16, 20, 24]; dropout: [0.1, 0.3, 0.5]; stride: [3, 5, 7];					

yielded a marginal reduction (~0.14%) in real-alarm recall, while simultaneously improved precision by 1.56%, F_2-score by 0.24%, and overall accuracy by 0.22% for the true early warning class. These results suggested that enhancing the correlation between rainfall and displacement contributed to improved model accuracy. Conversely, using only the masked self-attention module improved all metrics compared to the baseline, with gains ranging from 0.04% to 1.69%. This underscored the importance of enforcing temporal ordering in the attention mechanism for this task. Notably, the combined modules that formed the complete PatchDC model achieved the highest accuracy across the board, and showed a substantial 3.62% improvement in precision of true early warning over the baseline and demonstrated a synergistic effect between the two modifications.

Table 6 Accuracy results of Ablation Study for True early Warning recognition

	Precision of true early warning	Recall of true early warning	F_2-score of true early warning	Overall accuracy	AUPR
PatchTST	85.62±1.80	92.95±1.24	91.37±0.63	97.36±0.16	95.58±0.22
PatchTST＋Hybrid Feature Processing mechanism	87.18±1.12	92.81±1.82	91.61±1.39	97.58±0.22	96.06±0.43
PatchTST＋ masked self-attention module	87.31±1.12	92.95±2.06	91.75±1.63	97.61±0.26	95.62±0.48
PatchDC	89.24±2.43	93.38±0.95	92.50±0.49	97.91±0.28	96.35±0.54

https://doi.org/10.5194/egusphere-2026-2352
Preprint. Discussion started: 12 May 2026
© Author(s) 2026. CC BY 4.0 License.

6.3 Interpretability Analysis

To enhance model transparency and facilitate trust in critical decision-making scenarios, it was essential to illustrate the interpretability mechanism. PatchDC inherited the PatchTST backbone and processed input features largely independently through its initial layers. Its final classification layer used a Linear layer to analyze the feature representations from all components and output the prediction results. Therefore, the weight parameters of this Linear layer can indirectly reflect the importance assigned by the model to different input features for predicting the true early warning category.

(1) Feature Importance by average weight parameters

Fig. 9 displays the average weight parameters from the Linear layer associated with the true early warning class for each input feature, obtained by averaging over 25 independent training runs (5-fold cross-validation × 5 repetitions). The results indicated that the top three most important features were, in order: RDS (High-Resolution Raw Displacement Sequence), HMD (Hourly Median Displacement), and RMR (Relative Max Rainfall Intensity). This result aligned closely with the mechanism by which human experts assess warnings: recent high-resolution or long-term displacement data exhibiting a significant gradual increase trend served as the core basis for determining normal warnings; when displacement underwent significant sudden changes but the gradual trend was not evident, relative maximum rainfall intensity (RMR) became a critical auxiliary factor for assessment. Additionally, the fused features generated by the Hybrid Feature Processing mechanism (Fig. 9) ranked fourth in importance, further validating the effectiveness and contribution of this mechanism to the model's decision-making process.

(2) Synergistic Relationship of Feature Importance by weight parameters

Fig. 9 also reveals that the standard deviations of the feature weight parameters were generally large, indicating variability across training runs. This may be attributed to strong synergistic or compensatory relationships among the features regarding their importance for predicting the target variable. To investigate this, pairwise synergistic relationships between feature importances (Fig. 10) were analyzed through correlation analysis using the Pearson correlation coefficient, calculated from the weight parameters of the True Early Warning category across all 25 runs. The Spearman correlation coefficient is commonly used to assess multicollinearity between variables, with an absolute value exceeding 0.7 typically indicating a strong linear correlation (Passman et al., 2011). Applying this threshold to the pairwise correlations of feature importances, all absolute correlation values were found to be below 0.7, suggesting weak linear synergistic relationships between all features in terms of their importance weights.

(3) Feature Importance Stability Analysis and Validation by weight ranking

Given the verified weak linear synergistic relationships between features, the substantial standard deviation of feature weight parameters was more likely attributable to the following two factors. Firstly, data heterogeneity was a salient concern, characterized by the presence of distinct temporal feature patterns in training data from different folds. Secondly, model heterogeneity played a role: the optimal configuration (e.g., patch length) was independently determined through hyperparameter search during each training iteration, leading to differences in model structure and feature encoding

https://doi.org/10.5194/egusphere-2026-2352
Preprint. Discussion started: 12 May 2026
© Author(s) 2026. CC BY 4.0 License.

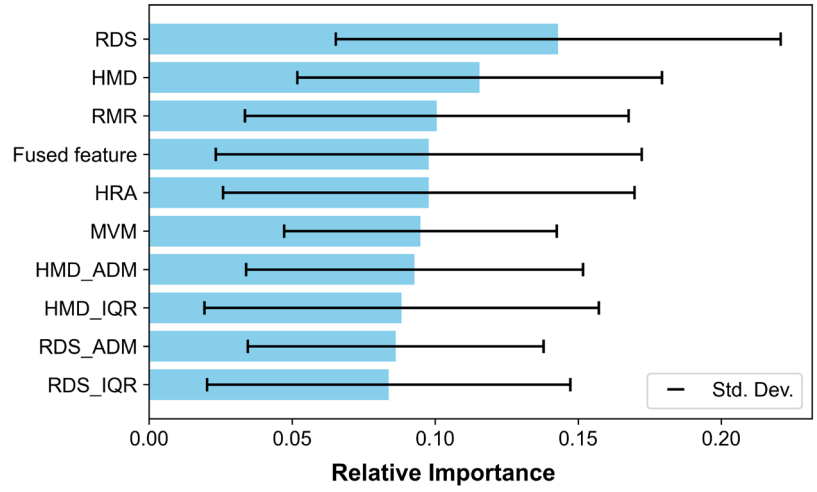


Figure 9 Feature Importance Based on Average weights for True Early Warning in the PatchDC Model

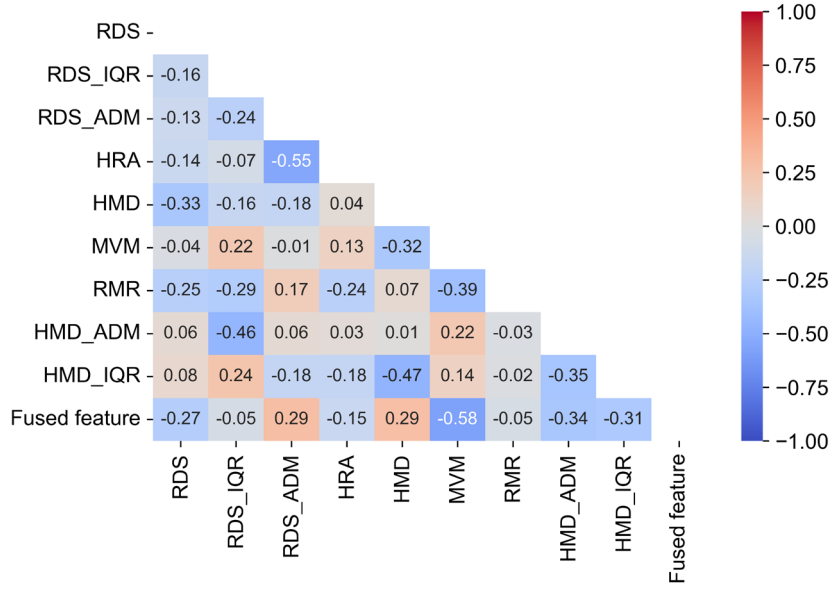


Figure 10 Correlation Analysis of Linear Feature Weights for True Early Warning in the PatchDC Model

semantics, which in turn affected the stability of the final linear layer weights. To further assess the reliability of the feature importance rankings despite this variability, the different rank of each feature weight across 25 training iterations was calculated. As shown in Table 7, RDS, HMD, and RMR are consistently the top-ranked features, with mean rankings of 4.24, 4.72, and 5.36, respectively; this indicated that the feature importance ranking aligned with the average weight-based ranking, confirming that these three were the most critical features. The mean ranking of the fused feature decreased to sixth place.

https://doi.org/10.5194/egusphere-2026-2352
Preprint. Discussion started: 12 May 2026
© Author(s) 2026. CC BY 4.0 License.

However, in the top 1 ranking and the top 3 proportion statistics, the fused feature ranked tied for third (12% of runs) and fourth (32% of runs) respectively, providing further evidence validating the contribution of the hybrid feature processing mechanism, even if its average rank was lower due to higher variability.

Table 7 Stability Analysis of Feature Importance Rankings Across 25 Training Iterations for True Early Warning

Features	Mean rank	Top1 count	Top3 count
RDS	4.24	7	13
HMD	4.72	3	9
RMR	5.36	4	7
MVM	5.44	2	6
HRA	5.48	3	11
Fused feature	5.6	3	8
RDS_ADM	5.88	1	4
HMD_ADM	5.96	1	5
HMD_IQR	6.12	1	6
RDS_IQR	6.2	0	6

6.4 Analysis of missed true early warnings and model limitations

To identify the model's blind spots and enhance future early warning reliability, a detailed analysis of false negatives (missed true warnings) was critical. Firstly, cases that were undetected (predicted as false) in ≥3 out of the 5 independent training runs were selected for in-depth analysis. A total of eight such true early warning instances were identified, distributed across only two sensors. Given the similarity in causes within each sensor, three representative missed cases were selected from each sensor and comparatively analyzed with correctly predicted warning cases from the same sensors.

(1) False Negative Analysis for Sensor 1 (Fig. 11 a-d, where d is a correctly predicted sample) was as follows. As illustrated in Fig. 11a, clear reverse displacement (contraction) is evident at the sensor's monitoring curve. This phenomenon was likely the result of continuous tensile cracking and subsidence at the landslide rear, as evidenced by the widening cracks observed on the village road. The landslide mass, which was rich in clay, may have encountered resistant strata, such as bedrock, or topographic obstruction downslope, resulting in temporary thrust uplift at the front (toe). It was important to note that such distinct reverse displacement patterns were never observed in the positive training samples available for the model, thereby hindering its ability to recognize them. As illustrated in Fig. 11b, the recent displacement (RDS feature) exhibited a positive (extensional) trend leading to the warning. However, the early reverse displacement significantly distorted the longer-term HMD feature, disrupting the overall progressive upward trend that the model learned to associate with true warnings. This reverse displacement signal of HMD was likely to have misled the model, resulting in a very low average positive prediction probability of 0.06 across the five runs. As shown in Fig. 11c, both displacement curves (RDS and HMD) exhibited a more pronounced recent upward trend than those in Fig. 11b, indicating continued landsliding. Consequently, the model's average

https://doi.org/10.5194/egusphere-2026-2352
Preprint. Discussion started: 12 May 2026
© Author(s) 2026. CC BY 4.0 License.

positive prediction probability increased from 0.06 to 0.22. As displacement progressed further to a state that depicted in Fig. 11d (a state that was correctly predicted), the average probability increased to 0.53. This finding aligned with the results of the interpretability analysis, suggesting that the model placed a premium on RDS features and particularly the recent upward trend in crackmeter displacement. However, Fig. 11c demonstrates that earlier anomalous displacement (reverse displacement) still exerted a significant interfering effect on the model's overall prediction confidence.

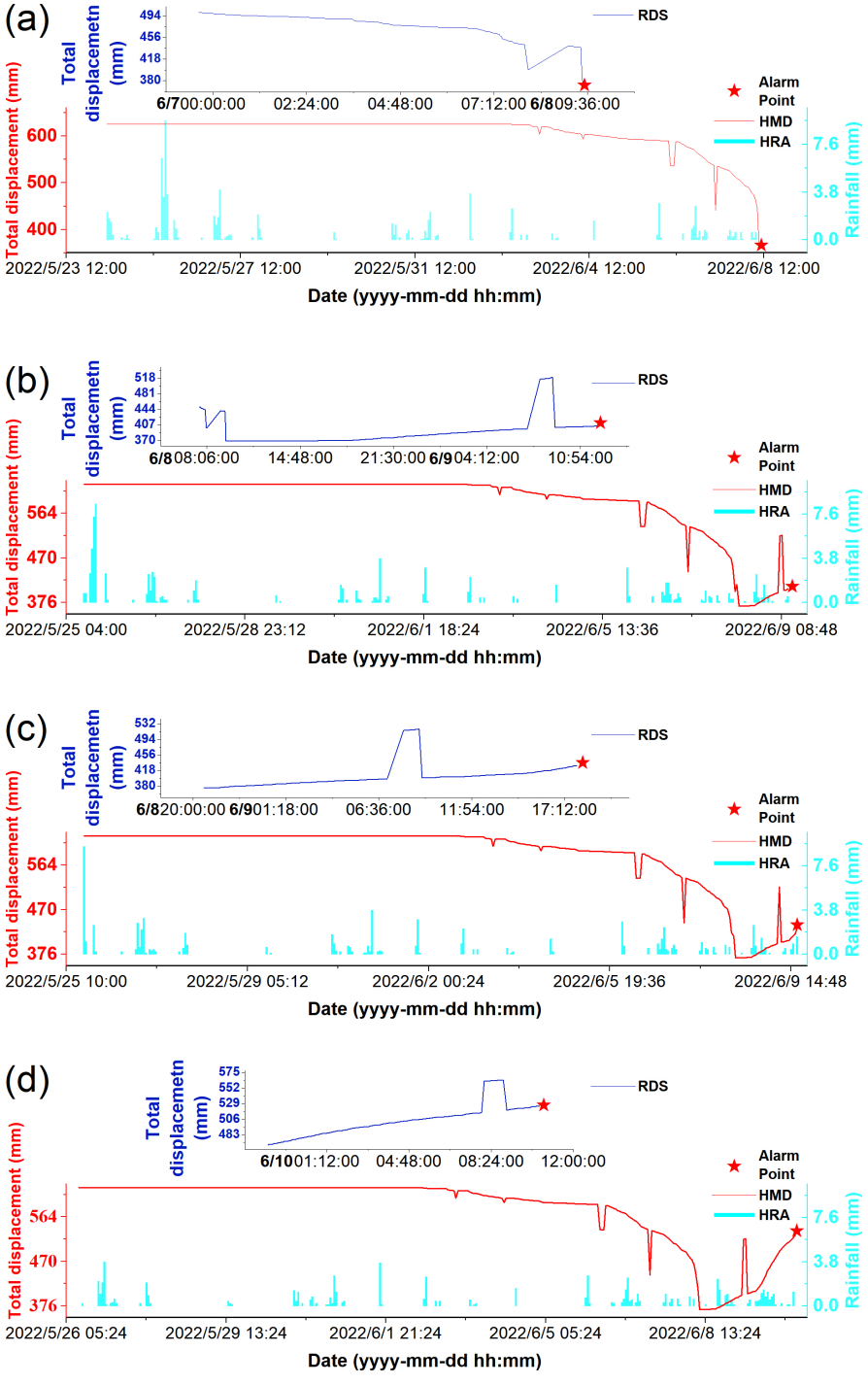
(2) False Negative Analysis for Sensor 2 (Fig. 12a-d, where a is a correctly predicted sample) was as follows. Missed detections for this sensor may be related to the recent displacement record of the crackmeter, which exhibited high-amplitude noise that lasted for a certain amount of time. The amplitude of these noises significantly exceeded the historical range of true displacement, potentially leading the model to misinterpret the overall trend as an increase followed by a decrease, rather than a continuous upward movement. Specifically, Fig. 12a (an early, correctly predicted case from the same sensor) shows smooth, noise-free displacement data with a clear trend, yielding a mean model prediction probability of 0.97. Following the appearance of significant abnormal noise (Fig. 12b and 12c), the mean prediction probability decreased substantially to 0.49 and 0.45, respectively, indicating model uncertainty. Fig. 12d show that the absence of a clear recent upward trend following the noise event caused the mean prediction probability to decline further to 0.10. These observations indicated that the model was insufficiently robust against high-amplitude noise that lasts for a certain amount of time. Compared with the correctly predicted case (Fig. 2), which also contained some anomaly, the noise amplitude in Fig. 2 did not exceed the historical true displacement range, which further illustrated the adverse impact of noise amplitude on model judgment. Furthermore, to verify that outlier interference was the primary cause of misclassification, the model was recalculated after manually removing these specific anomalies, resulting in correct predictions for all four cases shown in Fig. 12b-d.

Secondly, the reasons for other low-frequency false negatives (those occurring in ≤2 test runs) were analyzed as follows. In addition to inherent random fluctuations during training, these low-frequency false negatives often exhibited common patterns. As illustrated in Fig. 13, the model appears to underperform in predicting abrupt displacement surges that were triggered by prolonged periods of light-to-moderate rainfall, as opposed to short, intense downpours. Analytical results indicated that the model successfully predicted displacement surges associated with sudden, high-intensity rainfall events with a high degree of probability (e.g., Fig. 3). However, the model exhibited weaker discriminative ability for displacement surges triggered by sustained, light-to-moderate rainfall, as evidenced by probability estimates that oscillated near the decision boundary of 0.5. This phenomenon likely stemmed from the relative scarcity of such samples in the training dataset, resulting in the model's inability to firmly establish and generalize the mapping relationships between this specific rainfall pattern and its corresponding significant displacement surge.

In conclusion, the PatchDC model had several key limitations.

1) Limited reverse displacement detection capability: Insufficient representation of actual reverse displacement events in the warning signal samples impeded the model's ability to effectively monitor and correctly interpret such phenomena, which may indeed indicate genuine early warning stages of complex failure mechanics.

https://doi.org/10.5194/egusphere-2026-2352
Preprint. Discussion started: 12 May 2026
© Author(s) 2026. CC BY 4.0 License.



585 **Figure 11 Failure cases: the limited reverse displacement detection capability (correct prediction in (d))**

https://doi.org/10.5194/egusphere-2026-2352
Preprint. Discussion started: 12 May 2026
© Author(s) 2026. CC BY 4.0 License.

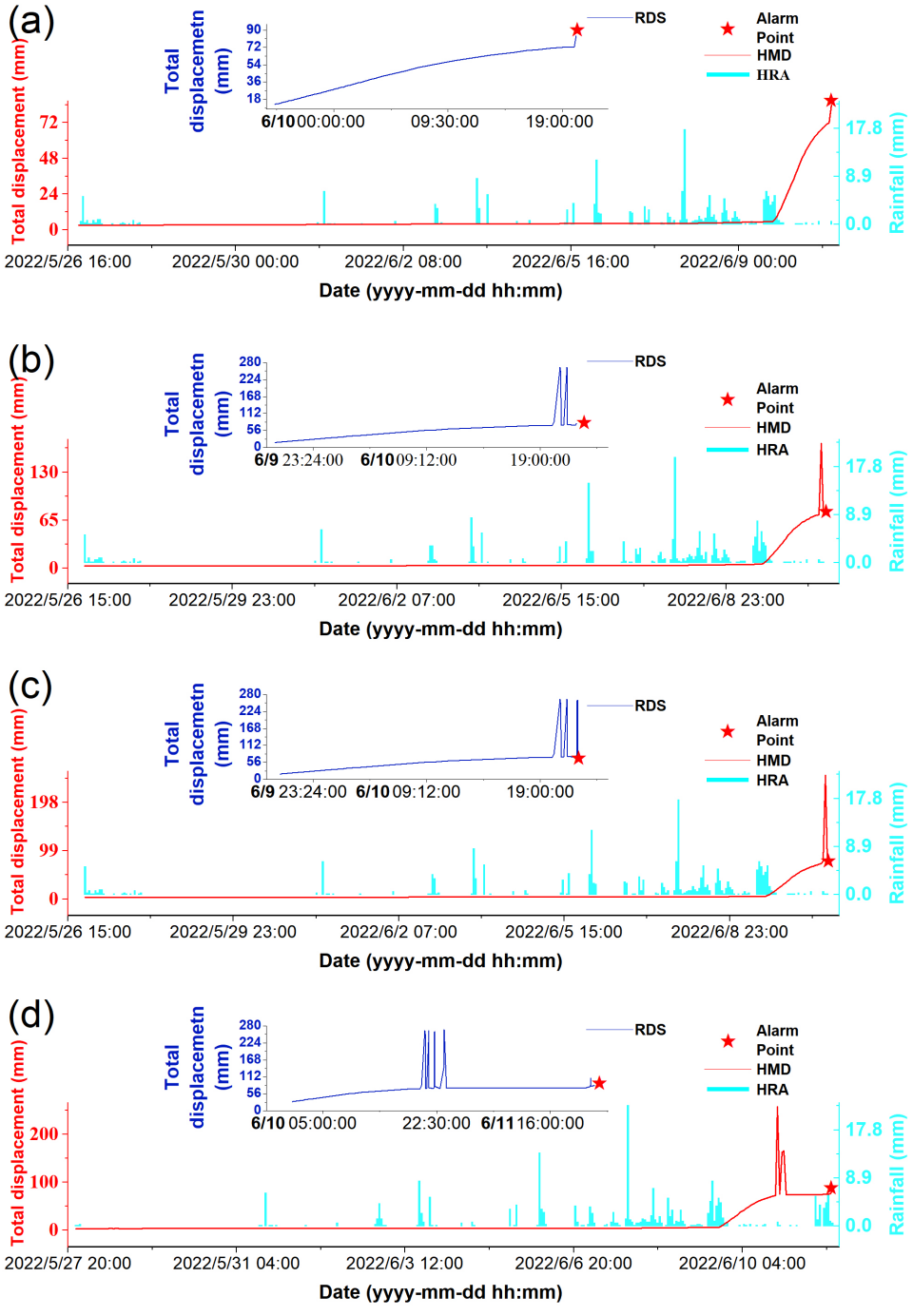
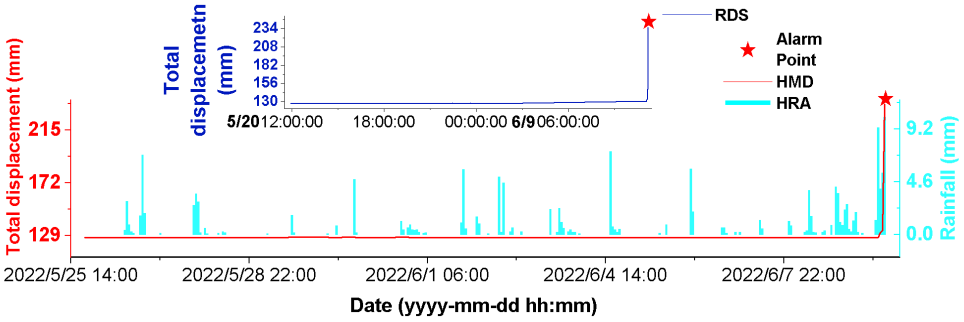


Figure 12 Failure Cases: the inadequate robustness against high-amplitude noise that lasts a certain amount of time (correct prediction in (a))

https://doi.org/10.5194/egusphere-2026-2352
Preprint. Discussion started: 12 May 2026
© Author(s) 2026. CC BY 4.0 License.



590 **Figure 13 Failure case: insensitivity to small-to-moderate rainfall-induced displacement surge**

2) Inadequate robustness against high-amplitude noise that lasts for a certain amount of time: The model may produce false negatives when there is a high frequency of large-amplitude anomalies in DSR time series data, especially when the amplitude exceeded the range of actual crack displacements.

595 3) Reduced sensitivity to small-to-moderate rainfall-induced displacement surges: The model exhibited weaker discriminative ability for abrupt displacement increases triggered by prolonged light-to-moderate rainfall compared to those triggered by short-duration rainstorms, likely due to underrepresented training examples for the former scenario.

4) Restricted applicability across triggering mechanisms: The model was primarily designed and trained for rainfall-induced geological hazards. Its effectiveness would be significantly diminished for events triggered by other mechanisms, such as

600 reservoir drawdown operations or anthropogenic engineering activities like excavation.

5) Dependency on paired instrumentation: Operational application of the model mandated the availability of concurrent data from both a crackmeter and a rain gauge. During data collection, frequent gaps occurred where crackmeter measurements lacked corresponding rainfall records (e.g., due to gauge malfunction) or where rain gauge data were missing during the critical warning period, effectively rendering the model inoperative for those events.

605 **6.5 Monitoring Process Improvement Based on Model Limitations**

Proactively refining monitoring workflows based on the identified model limitations is crucial for mitigating systemic vulnerabilities and preventing cascading errors.

In the case of limitation 2, if high-amplitude noise occurred for a certain amount of time and a warning point occurred near it, the noise itself would inevitably exceed the threshold and trigger a device warning. The model was then triggered to

610 recognize this noise as a false alarm. These noises could be excluded from the training data to mitigate noise interference. Even if the model mistakenly identifies them as genuine warnings, subsequent manual verification can correct the labeling. This enhances the model's generalization accuracy in future predictions, as demonstrated in Fig. 12 by the improved prediction accuracy after outlier removal.

https://doi.org/10.5194/egusphere-2026-2352
Preprint. Discussion started: 12 May 2026
© Author(s) 2026. CC BY 4.0 License.

For limitation 1 (actual reverse displacement), if a device issues warnings continuously over an extended period and the model persistently classifies them as false alarms, this persistently false alarms should trigger a mandatory manual verification process. Key areas of investigation should include potential device malfunctions and external interference factors (e.g., signal interference, or actual reverse displacement due to complex slope mechanics). On-site verification may be required to resolve the ambiguity.

For limitation 3 (low-frequency false negatives around the 0.5 decision threshold), the following strategies can be employed to further reduce the risk of missed detection. First, an ensemble learning strategy can be adopted, which involves training multiple independent models, such as five different instances. If any single model classifies an instance as a genuine warning, the final output is also classified as a genuine warning. This "one-vote positive" mechanism significantly improves the model's ability to detect anomalous samples. The second strategy is to lower the classification threshold from 0.5 to a lower level, such as 0.4. This will increase the recall (reduce missed alarms) but will also increase the number of false alarms, shifting the operating point on the PR curve.

7 Conclusion

Addressing the critical challenges of high false-alarm rates in crackmeter-based monitoring, substantial manual verification burdens, and difficulties in achieving reliable automated early warning for slope failures, this study constructed a specialized dataset combining crackmeter and rain-gauge measurements from different small-scale slope failure monitoring scenarios for real-time early warning of rainfall-induced small-scale slope failures. The proposed PatchDC model enabled real-time discrimination between true and false alarms across multiple crackmeter deployment scenarios. The optimally configured model achieved an F_2-score of 92.50%, demonstrating the efficacy of its dual-branch architecture for complementary long- and short-term multiscale feature extraction. PatchDC integrated a Hybrid Feature Interaction mechanism with a Temporally-aware Transformer encoder, outperforming the second-best deep learning baseline by 3.62% in Precision, 0.43% in Recall, 1.13% in F_2-score, and 0.55% in Overall Accuracy. This demonstrates its enhanced capability in capturing short-term directional spatiotemporal dependencies and learning explicit rainfall-displacement relationships. Interpretability analysis reveals that the top three contributing features in PatchDC are the raw displacement sequence (RDS), hourly median displacement (HMD), and relative maximum rainfall intensity (RMR), which aligns closely with the expert's prioritized criteria for authentic warning identification. To identify the model's blind spots and enhance operational reliability, a systematic case analysis identified five key limitations: 1) Limited identification capacity for true reverse-displacement warnings; 2) Insufficient robustness against high-amplitude noise that lasts a certain amount of time; 3) Reduced sensitivity to small-to-moder ate rainfall-induced displacement surges; 4) Primary applicability to rainfall-triggered geological hazards; and 5) Requirement for concurrent cleft and rain gauge data. An improved monitoring workflow was designed to alleviate some limitations, providing a more efficient tool for rainfall-induced small-scale slope failures that substantially reduces the expert verification burden. In addition, since all data originated from Fujian Province under relatively consistent monitoring

https://doi.org/10.5194/egusphere-2026-2352
Preprint. Discussion started: 12 May 2026
© Author(s) 2026. CC BY 4.0 License.

conditions, the current model's applicability is primarily validated within similar geological and instrumentation settings, while its broader generalization requires further exploration. Future work will focus on validating the model across diverse geological settings and sensor configurations, alongside integrating multi-source remote sensing and heterogeneous sensor data to enhance data availability and broaden applicability.

Code and data availability

Due to legal restrictions, the displacement and rainfall data used in this study cannot be publicly shared. PatchDC's code is available at https://github.com/ShubingOuyangcug/PatchDC.

Author contributions

SO developed the methodology and prepared the original draft of the manuscript. DL conceptualized the study, curated the data, acquired the funding, and supervised the research. ML carried out the investigation and contributed to data curation. FG developed the software and visualizations. XZ validated the findings and contributed to reviewing and editing the manuscript. Finally, YL and SW also contributed to reviewing and editing the manuscript.

Competing interests

The contact author has declared that none of the authors has any competing interests.

Acknowledgements

We gratefully acknowledge the Fujian Geologic Surveying and Mapping Institute for providing the crackmeter and rain-gauge measurements, along with the field-validated labels of true early warnings and false alarms used in this study. This work was supported by the National Key Research and Development Program of China [Grant number 2022YFB3904205] and the National Natural Science Foundation of China [Grant number 42301523].

https://doi.org/10.5194/egusphere-2026-2352
Preprint. Discussion started: 12 May 2026
© Author(s) 2026. CC BY 4.0 License.

Financial support

This work was supported by the National Key Research and Development Program of China [Grant number 2022YFB3904205] and the National Natural Science Foundation of China [Grant number 42301523].

References

675 Bai, S., Kolter, J. Z., and Koltun, V.: An empirical evaluation of generic convolutional and recurrent networks for sequence modeling, arXiv [preprint], arXiv:1803.01271, 2018.

Baroň, I. and Supper, R.: Application and reliability of techniques for landslide site investigation, monitoring and early warning–outcomes from a questionnaire study, Nat. Hazards Earth Syst. Sci., 13, 3157–3168, https://doi.org/10.5194/nhess-13-3157-2013, 2013.

680 Dahal, R. K. and Hasegawa, S.: Representative rainfall thresholds for landslides in the Nepal Himalaya, Geomorphology, 100, 429–443, https://doi.org/10.1016/j.geomorph.2008.01.014, 2008.

Deng, B., Xu, Q., Dong, X., Li, W., Wu, M., Ju, Y., and He, Q.: Automatic Method for Detecting Deformation Cracks in Landslides Based on Multidimensional Information Fusion, Remote Sens., 16, 4075, https://doi.org/10.3390/rs16214075, 2024.

685 Deng, L., Smith, A., Dixon, N., and Yuan, H.: Machine learning prediction of landslide deformation behaviour using acoustic emission and rainfall measurements, Eng. Geol., 293, 106315, https://doi.org/10.1016/j.enggeo.2021.106315, 2021.

Dou, H., Chen, Y., Sun, Y., and Guo, C.: Regional dynamic early warning model for rainfall-induced landslide in Fujian, China, Geomatics Nat. Hazards Risk, 15, 2435510, https://doi.org/10.1080/19475705.2024.2435510, 2024.

Ge, Q., Li, J., Wang, X., Deng, Y., Zhang, K., and Sun, H.: LiteTransNet: An interpretable approach for landslide
690 displacement prediction using transformer model with attention mechanism, Eng. Geol., 331, 107446, https://doi.org/10.1016/j.enggeo.2024.107446, 2024.

Guo, L., He, K., Liu, H., Meng, F., and Wang, X.: Physical prediction model of compound hydrodynamic unload-load response ratio and its application in reservoir colluvium landslide, J. Earth Sci., 35, 1304–1315, https://doi.org/10.1007/s12583-022-1662-9, 2024.

695 Guzzetti, F., Peruccacci, S., Rossi, M., and Stark, C. P.: The rainfall intensity–duration control of shallow landslides and debris flows: an update, Landslides, 5, 3–17, https://doi.org/10.1007/s10346-007-0112-1, 2008.

Hochreiter, S. and Schmidhuber, J.: Long short-term memory, Neural Comput., 9, 1735–1780, https://doi.org/10.1162/neco.1997.9.8.1735, 1997.

Jiang, Y., Zheng, L., Xu, Q., and Lu, Z.: Deformation mechanism-assisted deep learning architecture for predicting step-like
700 displacement of reservoir landslide, Int. J. Appl. Earth Obs. Geoinf., 133, 104121, https://doi.org/10.1016/j.jag.2024.104121, 2024.

https://doi.org/10.5194/egusphere-2026-2352
Preprint. Discussion started: 12 May 2026
© Author(s) 2026. CC BY 4.0 License.

Khan, A., Gupta, S., and Gupta, S. K.: Multi-hazard disaster studies: Monitoring, detection, recovery, and management, based on emerging technologies and optimal techniques, Int. J. Disaster Risk Reduct., 47, 101642, https://doi.org/10.1016/j.ijdrr.2020.101642, 2020.

Klimeš, J., Rowberry, M., Blahut, J., Briestenský, M., Hartvich, F., Košt'ák, B., Rybář, J., Stemberk, J., and Štěpančíková, P.: The monitoring of slow-moving landslides and assessment of stabilisation measures using an optical–mechanical crack gauge, Landslides, 9, 407–415, https://doi.org/10.1007/s10346-011-0306-4, 2012.

Lee, W. Y., Park, S. K., and Sung, H. H.: The optimal rainfall thresholds and probabilistic rainfall conditions for a landslide early warning system for Chuncheon, Republic of Korea, Landslides, 18, 1721–1739, https://doi.org/10.1007/s10346-020-01603-3, 2021.

Li, J., Wang, W., and Han, Z.: A variable weight combination model for prediction on landslide displacement using AR model, LSTM model, and SVM model: a case study of the Xinming landslide in China, Environ. Earth Sci., 80, 386, https://doi.org/10.1007/s12665-021-09696-2, 2021.

Li, S., Feng, W., Yi, X., Liu, K., Guo, C., Tang, X., and Wu, Z.: Clustered shallow landslides triggered by heavy rainfall in May 2022 in Wuping County, Fujian Province, China, Bull. Eng. Geol. Environ., 84, 257, https://doi.org/10.1007/s10064-025-04280-5, 2025.

Li, Z., Fang, L., Sun, X., and Peng, W.: 5G IoT-based geohazard monitoring and early warning system and its application, EURASIP J. Wirel. Commun. Netw., 2021, 160, https://doi.org/10.1186/s13638-021-02033-y, 2021.

Liu, Y., Hu, T., Zhang, H., Wu, H., Wang, S., Ma, L., and Long, M.: iTransformer: Inverted transformers are effective for time series forecasting, arXiv [preprint], arXiv:2310.06625, 2023.

Liu, Y., Long, J., Li, C., and Zhan, W.: Physics-informed data assimilation model for displacement prediction of hydrodynamic pressure-driven landslide, Comput. Geotech., 167, 106085, https://doi.org/10.1016/j.compgeo.2024.106085, 2024.

Luo, W., Dou, J., Fu, Y., Wang, X., He, Y., Ma, H., Wang, R., and Xing, K.: A novel hybrid LMD–ETS–TCN approach for predicting landslide displacement based on GPS time series analysis, Remote Sens., 15, 229, https://doi.org/10.3390/rs15010229, 2022.

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, 20, 2111–2129, https://doi.org/10.1007/s10346-023-02104-9, 2023.

Nie, Y., Nguyen, N. H., Sinthong, P., and Kalagnanam, J.: A time series is worth 64 words: Long-term forecasting with transformers, arXiv [preprint], arXiv:2211.14730, 2022.

Niu, P., Zhou, T., Wang, X., Sun, L., and Jin, R.: Attention as robust representation for time series forecasting, arXiv [preprint], arXiv:2402.05370, 2024.

Passman, M. A., McLafferty, R. B., Lentz, M. F., Nagre, S. B., Iafrati, M. D., Bohannon, W. T., Moore, C. M., Heller, J. A., Schneider, J. R., Lohr, J. M., et al.: Validation of venous clinical severity score (VCSS) with other venous severity

https://doi.org/10.5194/egusphere-2026-2352
Preprint. Discussion started: 12 May 2026
© Author(s) 2026. CC BY 4.0 License.

assessment tools from the American venous forum, national venous screening program, J. Vasc. Surg., 54, 2S–9S, https://doi.org/10.1016/j.jvs.2011.05.117, 2011.

Peng, S., Ding, Y., Liu, W., and Li, Z.: 1 km monthly temperature and precipitation dataset for China from 1901 to 2017, Earth Syst. Sci. Data, 11, 1931–1946, https://doi.org/10.5194/essd-11-1931-2019, 2019.

Qiu, H., Chen, X., Feng, P., Wang, R., Hu, W., Zhang, L., and Pasuto, A.: Advancing predictive accuracy of shallow landslide using strategic data augmentation, J. Rock Mech. Geotech. Eng., 17, 4273–4287, https://doi.org/10.1016/j.jrmge.2024.09.010, 2025.

Quansah, J. E., Engel, B., and Rochon, G. L.: Early warning systems: a review, J. Terr. Obs., 2, 5, https://docs.lib.purdue.edu/cgi/viewcontent.cgi?article=1072&context=jto, 2010.

Sharifi, S., Hendry, M., Macciotta, R., and Evans, T.: Evaluation of filtering methods for use on high frequency measurements of landslide displacements, Nat. Hazards Earth Syst. Sci. Discuss., 2021, 1–40, https://doi.org/10.5194/nhess-22-411-2022, 2021.

Sharifi, S., Macciotta, R., and Hendry, M. T.: A reliability evaluation of four landslide failure forecasting methods in real-time monitoring applications, Landslides, 21, 2545–2556, https://doi.org/10.1007/s10346-024-02293-x, 2024.

Sharifi, S., Macciotta, R., and Hendry, M. T.: Algorithms to enhance detection of landslide acceleration moment and time-to-failure forecast using time-series displacements, Eng. Geol., 309, 106832, https://doi.org/10.1016/j.enggeo.2022.106832, 2022.

Shehadeh, A., Alshboul, O., and Almasabha, G.: Slope displacement detection in construction: An automated management algorithm for disaster prevention, Expert Syst. Appl., 237, 121505, https://doi.org/10.1016/j.eswa.2023.121505, 2024.

Sun, X., Su, Y., Yang, C., Tan, J., and Liu, D.: Research on Slope Early Warning and Displacement Prediction Based on Multifractal Characterization, Fractal Fract., 8, 522, https://doi.org/10.3390/fractalfract8090522, 2024.

Zeng, T., Jiang, H., Liu, Q., and Yin, K.: Landslide displacement prediction based on Variational mode decomposition and MIC-GWO-LSTM model, Stoch. Environ. Res. Risk Assess., 36, 1353–1372, https://doi.org/10.1007/s00477-021-02145-3, 2022.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I.: Attention is all you need, Adv. Neural Inf. Process. Syst., 30, https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf, 2017.

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

Wang, Z., Tang, J., Hou, S., Wang, Y., Zhang, A., Wang, J., Wang, W., Feng, Z., Li, A., and Han, B.: Landslide displacement prediction from on-site deformation data based on time series ARIMA model, Front. Environ. Sci., 11, 1249743, https://doi.org/10.3389/fenvs.2023.1249743, 2023.

https://doi.org/10.5194/egusphere-2026-2352
Preprint. Discussion started: 12 May 2026
© Author(s) 2026. CC BY 4.0 License.

Wu, D., Zhou, B., and Zimin, M.: Prediction of landslide displacement based on the CA-stacked transformer model, Alexandria Eng. J., 124, 389–403, https://doi.org/10.1016/j.aej.2025.03.140, 2025.

Wu, H., Hu, T., Liu, Y., Zhou, H., Wang, J., and Long, M.: Timesnet: Temporal 2d-variation modeling for general time series analysis, arXiv [preprint], arXiv:2210.02186, 2022.

Wu, Y. and Lan, H.: Landslide Analyst—a landslide propagation model considering block size heterogeneity, Landslides, 16, 1107–1120, https://doi.org/10.1007/s10346-019-01154-2, 2019.

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

Yang, B., Yin, K., Lacasse, S., and Liu, Z.: Time series analysis and long short-term memory neural network to predict landslide displacement, Landslides, 16, 677–694, https://doi.org/10.1007/s10346-018-01127-x, 2019.

Yang, J., Huang, Z., Jian, W., and Robledo, L. F.: Landslide displacement prediction by using Bayesian optimization–temporal convolutional networks, Acta Geotech., 19, 4947–4965, https://doi.org/10.1007/s11440-023-02205-8, 2024.

Zeng, A., Chen, M., Zhang, L., and Xu, Q.: Are transformers effective for time series forecasting?, in: Proceedings of the AAAI Conference on Artificial Intelligence, 11121–11128, https://doi.org/10.1609/aaai.v37i9.26317, 2023.

Zhang, J., Tang, H., Li, C., Gong, W., Zhou, B., and Zhang, Y.: Deformation stage division and early warning of landslides based on the statistical characteristics of landslide kinematic features, Landslides, 21, 717–735, https://doi.org/10.1007/s10346-023-02192-7, 2024.

Zhuang, Y., Xing, A., Jiang, Y., Sun, Q., Yan, J., and Zhang, Y.: Typhoon, rainfall and trees jointly cause landslides in coastal regions, Eng. Geol., 298, 106561, https://doi.org/10.1016/j.enggeo.2022.106561, 2022.