August 11, 2026

To: The Editor-in-Chief, The Associate Editor, And the Reviewers
Natural Hazards and Earth System Sciences

Title: Reducing False Alarms in Crackmeter-Based Early Warning of Small-Scale Slope Deformation via Deep Learning-Driven Asynchronous Displacement–Rainfall Data Fusion

Authors: Shubing Ouyang, Daichao Li*, Minjiang Liu, Fengjian Ge, Xia Zheng, Yuan Li, and Sheng Wu

Dear Timothy Tiggeloven and Reviewers:

Thank you very much for your response. We thank the editor and reviewers for the time and effort that they put into reviewing the previous version of the manuscript. Following the reviewers' comments, we have carefully modified and improved our manuscript. We hope these revisions now resolve the problems and uncertainties pointed out by the reviewers.

A point-by-point response to the reviewers' comments is provided in the following pages.

For clarity, we note at the outset that the numerical results in the revised manuscript differ from those reported previously because the experimental protocol was substantially revised. Specifically, (1) the dataset was repartitioned to enforce strict sensor-level isolation among the training, validation, and test sets, including within the five-fold cross-validation; (2) all normalization parameters were estimated exclusively from the corresponding training subset and then applied unchanged to the validation and test sets; (3) the positive-class weight was recalculated for each fold based on its training-set class distribution; and (4) MVM was removed as an independent input channel and retained only for fused-feature construction. All models were retrained under this revised, leakage-controlled protocol, and the corresponding results, tables, figures, and conclusions were updated accordingly. Therefore, all conclusions in the revised manuscript are based on the newly generated results rather than those reported in the previous version.

We sincerely hope that our manuscript is now suitable for publication in Natural Hazards and Earth System Sciences.

Thank you for your assistance and we look forward to hearing from you.

Yours sincerely,

Daichao Li and Shubing Ouyang
The Academy of Digital China (Fujian)
Fuzhou University,
Fuzhou, China
Email: lidc@fzu.edu.cn; 394666@fzu.edu.cn

PART I. REVIEWER 1

The manuscript addresses an important and relevant topic within landslide early warning systems, and the machine-learning/data-fusion methodology appears technically rigorous and competently validated.

1.However, the manuscript currently treats landslide early warning primarily as a signal-classification problem, without proper consideration of the geotechnical, geological and hydromechanical aspects.

Response:

Sensor locations were selected by geological experts following field investigations and were placed across identified rear tension cracks, with consideration of slope geometry, lithology, drainage conditions, and interpreted failure mechanisms. These site-specific factors also informed threshold setting and expert validation. However, detailed site-characterization data were not systematically collected for this study and therefore could not be included as model inputs, which represents a limitation of the present work. Future work will integrate geological and environmental information more explicitly, including the use of landslide-susceptibility or hazard maps derived from these factors as spatial priors for model prediction.

The added text is as follows:

Page 4 (The end of introduction): To address this practical challenge, this study focuses on distinguishing genuine alerts from noise-related false alarms after crackmeter thresholds are exceeded in an operational slope-monitoring network in Fujian Province, China. In this network, crackmeters are installed across rear tension cracks identified through field investigations at sites affected predominantly by small-scale landslides or localized collapses. Geological experts establish a site-specific four-level threshold scheme for each crackmeter based on local geological and environmental conditions. Exceedance of a threshold generates a preliminary device alert rather than confirming that slope failure is imminent. Experts must therefore determine whether the recorded displacement represents genuine slope deformation or interference caused by sensor noise, human or animal disturbance, or other non-slope-related factors. This discrimination must be performed promptly because deformation at some monitored small-scale slope may accelerate rapidly, while even lower-level alerts may remain relevant when continued or forecast rainfall could further promote crack opening and slope movement.

The added text is as follows:

Page 37 (The end of Conclusion)：Future work will integrate geological and environmental information more explicitly, including the use of landslide-susceptibility or hazard maps derived from these factors as spatial priors for model prediction.

2.The title, abstract and introduction should more clearly define the scope and contribution of the study. The novelty appears to lie mainly in the dataset manipulation and ML workflow.

Response:

Thank you for this valuable suggestion. We apologize that the scope and contributions of the study were not sufficiently clear in the previous manuscript. To more clearly define the study as a post-threshold true–false warning discrimination task rather than a complete slope-stability or failure-prediction framework, we have revised the Title, Abstract, and Introduction. We further clarified that crackmeters were installed across field-identified rear tension cracks and that site-specific warning thresholds were established by geological experts based on local geological and environmental conditions. The stated novelty was refocused on the field-verified dataset and multi-rate displacement–rainfall fusion framework.

The revised Title, Abstract, and Introduction are as follows (The main revised parts are marked in blue):

(1) The title: Reducing False Alarms in Crackmeter-Based Early Warning of Small-Scale Slope Deformation via Deep Learning-Driven Asynchronous Displacement–Rainfall Data Fusion

(2) The abstract: Real-time crackmeter-based deformation monitoring systems for small-scale rainfall-affected slopes suffer from high false alarm rates. Across multi-slope monitoring settings, threshold-based alerts are vulnerable to noise, while conventional denoising and prediction-based validation methods remain limited in automated warning discrimination and cross-site generalization. To address this bottleneck, a labeled dataset and a data-driven binary classification framework were developed to distinguish true early warnings from false alarms among threshold-triggered sensor alerts. A multi-source dataset of crack-displacement and rainfall observations was constructed from monitored slopes in Fujian Province, China, with genuine and false alerts labelled based on field investigations and geological expert assessment. A patch-based dual-branch temporally-aware Transformer model was developed to explicitly address asynchronous multi-rate data fusion, strict temporal ordering between antecedent rainfall and subsequent 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 90.32%, a recall of 89.38%, and an F_2-score of 89.54% in identifying true early warnings. Interpretability analysis reveals that the model's decisions are primarily driven by localized displacement trends and the relationship between rainfall and displacement, 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. However, the present validation is limited to the Fujian rainfall-triggered monitoring context, and application to other geological or climatic settings requires further external validation and local adaptation.

(3) The Introduction, including the statement of contributions:

Page 4 (The end of the introduction): To address this practical challenge, this study focuses on distinguishing genuine alerts from noise-related false alarms after crackmeter thresholds are exceeded in an operational slope-monitoring network in Fujian Province, China. In this network, crackmeters are installed across rear tension cracks identified through field investigations at sites affected predominantly by small-scale landslides or localized collapses. Geological experts establish a site-specific four-level threshold scheme for each crackmeter based on local geological and environmental conditions. Exceedance of a threshold generates a preliminary device alert rather than confirming that slope failure is imminent. Experts must therefore determine whether the recorded displacement represents genuine slope deformation or interference caused by sensor noise, human or animal disturbance, or other non-slope-related factors. This discrimination must be performed promptly because deformation at some monitored small-scale slope may accelerate rapidly, while even lower-level alerts may remain relevant when continued or forecast rainfall could further promote crack opening and slope movement. To support this process, 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 temporal ordering between antecedent rainfall and subsequent 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. It should be noted that the operational monitoring network comprised 429 crackmeters, whereas field-confirmed true-warning samples were obtained from 52 sensors. Accordingly, the present validation is limited to the rainfall-triggered monitoring context in Fujian Province, and the resulting model should not be assumed to generalize directly to other geological, climatic, or monitoring settings.

The main contributions of this study are as follows: 1) A field-verified multi-source dataset was constructed for true–false warning discrimination across multiple monitored slopes, integrating crackmeter displacement and rainfall observations following threshold-triggered alerts. 2) A temporally aware dual-branch Transformer framework was developed to integrate asynchronous multi-rate rainfall and displacement data for automated warning discrimination. The framework was evaluated under strict sensor-level isolation within the Fujian monitoring network, demonstrating reduced false alarms while retaining sensitivity to field-confirmed true warnings from previously unseen sensors.

3. The term “small-scale” is ambiguous and should be clarified.

Response:

Thank you for pointing out the ambiguity in the term “small-scale.” We apologize for not clearly defining this term in the original manuscript. We have now added a quantitative definition at its first occurrence.

The added text is as follows:

Page 2 (the first paragraph of the introduction): According to the classification of the Ministry of Natural Resources of China, small-scale landslides are defined as those with a volume of less than 100,000 m^3, whereas small-scale collapses are defined as those with a volume of less than 10,000 m^3.

4. A concern is the reliance on crack meters as the principal monitoring approach for shallow landslide warning. Crack opening is not necessarily representative of bulk slope deformation or impending instability, and the manuscript does not sufficiently justify this monitoring philosophy. Later sections imply that this approach has already been widely deployed across Fujian Province, and that the AI framework is intended to mitigate resulting false alarms. If so, this context should be introduced much earlier.

The manuscript also gives insufficient consideration to instrumentation placement and site characterisation. Sensor location relative to slope geometry, drainage, lithology and likely failure mechanism is critical to reliable interpretation.

Response:

Thank you for this valuable suggestion. We apologize for not clearly describing the operational monitoring framework, which may have caused confusion regarding the role of crackmeters and the logic of the proposed model. First, crackmeters are not installed across arbitrary surface cracks; they are deployed across rear tension cracks identified through field investigations at sites affected predominantly by small-scale landslides or localized collapses. Second, the warning threshold is not fixed or uniformly applied across all sites. Instead, geological experts establish a site-specific four-level threshold scheme according to the local geological and environmental conditions. Exceedance of a threshold generates a preliminary device alert rather than a direct prediction of imminent slope failure. Geological experts must then determine whether the recorded displacement reflects genuine slope deformation or noise-related interference. At these predominantly small-scale sites, deformation may accelerate rapidly once slope instability develops, leaving a limited time window for manual verification and response. Moreover, lower-level threshold exceedances may still provide meaningful early-warning information, because continued or forecast heavy rainfall may further amplify crack opening and increase the likelihood of subsequent slope movement.

The added text is as follows:

Page 4 (The end of introduction): To address this practical challenge, this study focuses on distinguishing genuine alerts from noise-related false alarms after crackmeter thresholds are exceeded in an operational slope-monitoring network in Fujian Province, China. In this network, crackmeters are installed across rear tension cracks identified through field investigations at sites affected predominantly by small-scale landslides or localized collapses. Geological experts establish a site-specific four-level threshold scheme for each crackmeter based on local geological and environmental conditions. Exceedance of a threshold generates a preliminary device alert rather than confirming that slope failure is imminent. Experts must therefore determine whether the recorded displacement represents genuine slope deformation or interference caused by sensor noise, human or animal disturbance, or other non-slope-related factors. This discrimination must be performed promptly because deformation at some monitored small-scale slope may accelerate rapidly, while even lower-level alerts may remain relevant when continued or forecast rainfall could further promote crack opening and slope movement.

5. The definition of “false alarms” remains unclear throughout the manuscript and requires substantial clarification. It is currently uncertain whether false alarms refer to:
erroneous sensor readings,
genuine displacement that did not progress to failure,
conservative warnings,
or deformation unrelated to landslide processes.

These are fundamentally different phenomena with different implications for warning-system performance. Similar clarification is needed for Figure 4 and the “triple-blind review” discussed at line 205.
Figures 11–13 also require clarification. It is not clear whether the timing of actual landslide occurrence is shown relative to the issued alarms.

Response:

Thank you for this valuable suggestion. Apologies for the unclear definition of “false alarms” in the original text.

The following supplements the category definitions in detail:

Page 8-9 (3.1 Problem Definition and Methodological Framework): Class 1 (true early warning) referred to an alert for which the recorded displacement change was subsequently confirmed by geological experts, based on professional assessment and field evidence, as genuine physical opening or displacement of the monitored rear tension crack associated with slope activity. A positive label did not require that complete slope failure subsequently occurred.

Class 0 (false alert) referred to an alert for which the apparent threshold exceedance was not confirmed as genuine and reliable deformation of the monitored rear tension crack. Such alerts mainly included abrupt data spikes, abnormal oscillations, sensor or communication malfunctions, accidental mechanical interference caused by animals or human contact, and other anomalous responses unrelated to actual rear-crack deformation.

The caption of Figure 4 can also be updated accordingly to:

Figure 4. Example of a false crackmeter alert caused by an apparent threshold exceedance that was not validated as genuine rear-crack deformation. Possible causes of this type of response include instrumental anomalies, installation disturbance, and accidental contact by animals or humans.

The following supplements about the "triple-blind review" discussed in detail:

Page 10 (3.2 Dataset Construction): Specifically, 173 temporally adjacent instances underwent independent blinded review by three geological experts, and only those unanimously identified as genuine crack-opening events were included in the dataset.

The captions of Figures 11–13 have also been supplemented with the following notes:

The figures show threshold-based true early warnings but do not indicate that slope failure occurred immediately.

6. The statement that landslides are "most strongly associated with precipitation effects within a 6–10 day window" is overly broad and insufficiently justified. Slope response depends strongly on material type, permeability, drainage conditions and failure depth. However, the manuscript provides very limited information regarding geology, geomaterials or failure mechanisms.

Response:

Thank you for these important comments. We apologize that our previous statement was inaccurate.

This is hereby corrected as follows:

Page 10 (3.2 Dataset Construction): Previous studies have demonstrated that antecedent rainfall and delayed slope responses may extend over several days to just over a week, although the relevant timescale varies among sites (Dahal and Hasegawa, 2008; Dou et al., 2024; Lee et al., 2021). Such variability is controlled by geomaterial properties, permeability, drainage conditions, antecedent moisture, failure depth, and failure mechanism. Therefore, this study did not assume a universal rainfall-response period. Instead, a 360-h retrospective window preceding each device alert was adopted as a fixed operational input interval. The 15-day duration was selected to encompass the multi-day rainfall effects reported in previous studies while providing additional coverage for delayed responses across heterogeneous monitoring sites. Longer windows of 30 or 60 days were not used because they may include rainfall episodes less directly related to the current alert and dilute the recent rainfall–displacement relationship. Thus, the 15-day window represents a pragmatic modelling choice rather than a universally optimal hydrological response period.

7. The machine-learning workflow appears rigorous, but the manuscript does not compare the proposed framework against simpler or more conventional approaches that may already be sufficient to reduce false alarms. This raises the question of whether a highly sophisticated AI framework is being used to compensate for limitations in an oversimplified monitoring philosophy.

Response:

Thank you for this valuable comment. To address this concern, we added comparisons with a conventional Bayesian rainfall intensity–duration (I–D) threshold and three simple machine-learning models that do not use deep learning: logistic regression, a decision tree, and a random forest. These comparisons evaluate whether simpler approaches are sufficient for false-alarm reduction. The results show that although the conventional methods perform well in certain metrics, PatchDC achieves a substantially better balance between identifying true warnings and reducing false alarms.

We also note that the numerical results in the revised manuscript differ from those reported in the previous version because several methodological corrections and refinements were implemented during revision. First, the dataset was repartitioned to enforce **sensor-level isolation among the training, validation, and test sets**, and the same sensor-isolation constraint was maintained within the five cross-validation folds. Second, the normalization procedure was revised so that all normalization parameters were estimated **exclusively from the corresponding training subset** and then applied unchanged to the validation and test data, thereby preventing information leakage. Third, instead of using the previously fixed 10:1 class-weighting ratio, the positive-class weight in the loss function was recalculated separately for each fold according to the actual negative-to-positive ratio in that fold's training subset. Finally, the long-term branch of PatchDC was refined by removing MVM as an independent input channel; MVM is now used only within the Hybrid Feature Processing module for fused-feature construction. Accordingly, **all conclusions in the revised manuscript are based on the newly generated results under the revised and leakage-controlled experimental protocol**, rather than on the numerical values reported in the previous version.

The added text and table are as follows:

Page 19-20: 5.1 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); Logistic regression, a decision tree, and a random forest were selected as conventional supervised machine-learning baselines. These models used 23 physically interpretable predictors derived from the same 15-day pre-alert hourly rainfall–displacement sequence and 100-point high-frequency displacement sequence provided to PatchDC (Table 6). Missing values for the conventional machine-learning models were imputed using medians estimated only from the corresponding training fold, and logistic regression additionally used robust scaling fitted on that fold.

Among the conventional machine-learning baselines (Table 7), random forest provided strong false-alarm discrimination, with precision only 2.39 percentage points below PatchDC; however, its recall and F_2-score were lower by 24.54 and 21.10 percentage points, respectively. Decision tree offered a more balanced precision–recall relationship, although PatchDC still exceeded it by approximately 14 percentage points in precision, recall, and F_2-score. Logistic regression favored recall but generated more false positives, with deficits of 29.61 percentage points in precision and 14.54 in F_2-score relative to PatchDC. Finally, the rainfall-only Bayesian Intensity–Duration (I–D) threshold achieved recall 6.56 percentage points higher than PatchDC, reflecting a highly conservative warning strategy. Nevertheless, it was lower than PatchDC by 71.70 percentage points in precision, 37.15 in F_2-score, 43.10 in overall accuracy, and 65.11 in AUPR. This high false-alarm rate illustrates the limitation of relying solely on rainfall triggers without considering the observed kinematic response of the slope and further highlights the operational value of PatchDC's multi-source, multi-resolution feature-fusion framework.

Table 6 Definition of the handcrafted rainfall–displacement predictors used by the conventional machine-learning baselines

Feature group	Predictor	Calculation	Temporal scale	Number
Rainfall	Antecedent cumulative rainfall	Sum of the hourly rainfall amounts recorded within each pre-alert window.	Previous 24, 72, 168, and 360 h	4
	Maximum 1-h rainfall intensity	Maximum hourly rainfall amount during the 24 h immediately preceding the alert.	Previous 24 h	1
	Relative maximum rainfall intensity (RMR) is defined in the same manner as in Table 3			1
Hourly crack displacement	Net displacement change	Last minus first hourly displacement value within each pre-alert window (Dend - Dstart).	Previous 24, 72, 168, and 360 h	4
	Linear displacement trend	Slope of a least-squares linear fit between hourly displacement and observation order within each window.	Previous 24, 72, 168, and 360 h	4
	HMD interquartile range (HMD_IQR) is defined in the same manner as in Table 3			1
	HMD absolute displacement magnitude (HMD_ADM) is defined in the same manner as in Table 3			1
High-frequency crack displacement	Net displacement change	Last minus first displacement measurement among the most recent N chronologically ordered observations.	Most recent 10, 30, 60, and 100 observations	4
	Maximum absolute adjacent increment	Maximum absolute difference between two consecutive displacement measurements among the most recent 100 observations.	Most recent 100 observations	1
	RDS interquartile range (RDS_IQR) is defined in the same manner as in Table 3			1
	RDS absolute displacement magnitude (RDS_ADM) is defined in the same manner as in Table 3			1

Table 7. Accuracy results and hyperparameter search spaces of the evaluated 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	18.62±0.74	**95.94±0.35**	52.39±1.16	54.73±2.31	29.67±2.36
Logistic regression	Handcrafted predictors	60.71 ± 0.00	79.69 ± 0.00	75.00 ± 0.00	92.33 ± 0.00	71.21 ± 0.00
Decision tree	Handcrafted predictors	75.64 ± 0.66	75.16 ± 1.40	75.25 ± 1.02	94.76 ± 0.07	81.34 ± 1.05
Random forest	Handcrafted predictors	87.93 ± 0.51	64.84 ± 0.00	68.44 ± 0.06	95.30 ± 0.05	88.53 ± 0.08
TCN	Dual-branch	84.72±2.17	78.44±2.25	79.6±1.73	96.18±0.25	84.73±1.08
LSTM	Dual-branch	80.09±1.87	82.34±3.06	81.87±2.67	95.93±0.45	88.89±1.24
TimesNet	Dual-branch	85.85±1.81	84.69±3.81	84.87±2.78	96.86±0.17	92.17±1.39
DLinear	Dual-branch	67.74±1.13	78.28±3.55	75.88±2.55	93.69±0.19	73.2±1.58
PatchTST	Dual-branch	90.14±2.16	82.81±1.91	84.18±1.9	97.2±0.4	92.87±1.44
iTransformer	Dual-branch	80.03±5.36	84.22±5.7	83.16±3.54	96±0.47	86.1±1.8
PatchDC	Dual-branch	**90.32±2.64**	89.38±1.8	**89.54±1.09**	**97.83±0.22**	**94.78±0.56**
The hyperparameter search space	Bayesian I–D Threshold: IETD ∈ {6, 12, 24, 32} h; $I = \alpha D^{-\beta}$; $1/\alpha \sim U(0.001, 10)$, $\beta \sim U(0.1, 2.0)$; selected by mean validation F_2. Search sampling: 500 draws + 500 tuning steps, 2 chains; final refit: 2000 draws + 1000 tuning steps, 4 chains. Logistic regression: RobustScaler; class_weight = balanced; solver = liblinear; max_iter = 5000. Grid: C ∈ {0.001, 0.01, 0.1 , 1, 10, 100}; penalty ∈ {L1, L2} Decision tree: class_weight = balanced. Grid: max_depth ∈ {2, 3, 5, None}; min_samples_leaf ∈ {5, 10, 20} Random forest: n_estimators = 500; max_features = sqrt; class_weight = balanced_subsample; bootstrap = True. Grid: max_depth ∈ {4, 8, None}; min_samples_leaf ∈ {2, 5, 10} TCN: num_levels ∈ {8, 16, 32, 64}; dropout ∈ {0.1, 0.3, 0.5}; branch_channels ∈ {1×, 2×, 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}.					

8. The discussion should also engage more critically with operational geotechnical practice: warning thresholds should generally be site-specific and informed by local geology and failure mechanisms, rather than applied uniformly across many slopes. Warnings are based not only on threshold exceedance, but also on trends in behaviour, rates of change, and engineering judgement.

Response:

Thank you for raising this important point. We apologize that the threshold-setting procedure was not clearly explained in the original manuscript. A single uniform threshold is not applied to all monitored slopes. Rather, geological experts establish a fixed, site-specific threshold for each slope according to its local geological and monitoring conditions.

We have revised the relevant text to clarify this issue, as follows:

Page 5 (2 The Engineering Problem: Early Warning Characteristics and System Limitations): The operational monitoring system uses threshold-based trigger rules to support continuous and real-time monitoring, but the thresholds are not applied uniformly across all slopes. Geological experts establish site-specific, four-level thresholds for each crackmeter according to the local geological setting, slope geometry, drainage conditions, interpreted failure mechanism, and monitoring conditions.

9. The manuscript would benefit from discussion of the transferability of the proposed methodology beyond Fujian Province and to settings.

Response:

Thank you for this valuable suggestion. We agree that the transferability of the proposed methodology should be discussed more clearly.

The added text is as follows:

Page 30: 6.6 Transferability and limitations

The potentially transferable contribution of this study lies in its problem formulation and modelling framework rather than in the Fujian-trained model itself. Its broader relevance rests on two considerations. First, displacement at rainfall-triggered slope instability sites in other regions may also exhibit progressive development. Second, the relationship between rainfall and displacement should be explicitly considered when assessing whether crackmeter alerts reflect physically meaningful slope responses. The proposed framework addresses both aspects by integrating rainfall and crack-displacement observations at different temporal resolutions. However, it has currently been validated only at predominantly small-scale, rainfall-triggered landslide and collapse sites in Fujian Province, where crackmeters were installed across field-identified rear tension cracks. Its applicability under different geological, climatic, failure-mechanism, or monitoring conditions therefore remains to be established through locally labelled data, model retraining or recalibration, and independent external validation.