Responses to the First Reviewer's Comments

Pan Jiang, Zhengjing Ma*, Gang Mei**
School of Engineering and Technology
China University of Geosciences (Beijing)
gang.mei@cugb.edu.cn

Acknowledgement The authors would like to thank the editor and the reviewers for their comments.

Comment # 1:

1. The paper is well-organized and includes all the key elements I would expect from a review article on this topic. The topic itself (the use of deep learning in identifying landslides) is a worthwhile area to have a review on. The discussion of landslide mechanisms in Section 4 is quite detailed and interesting (although needs more discussion of how deep learning has been applied, see Weaknesses). The recommendations in Section 6, particularly regarding data fusion, feel well-motivated and salient. The figures are mostly helpful and illustrative overall (see specific comments below), and I think would be useful for a landslide researcher who wants to know more about deep learning. The review of previous literature is extensive and can point curious readers in the right direction.

Response:

- Dear reviewer, we appreciate your positive and encouraging evaluation of our manuscript! We are very encouraged by their assessment that the paper is well-organized, comprehensive, and a worthwhile contribution to the field. We have carefully considered all the comments and have revised the manuscript accordingly. The point-by-point responses are detailed below.
- *** Regarding the discussion of landslide mechanisms in Section 4**
- Following your constructive suggestion regarding Section 4, we have supplemented specific deep learning application cases for each major type of landslide. In particular, we have elaborated on how the strengths of different models align with the recognition requirements of specific landslide types. Please see **Comment #27**.
- *** Regarding the discussion of figures**
- Thank you for the positive recognition. Based on your specific suggestions, we have carefully considered and provided our replies. Please see **Comment #7 and Comment #31** for details.
-
- Once again, we express our sincere gratitude to the reviewer for their valuable time and constructive comments, which have undoubtedly improved the quality of our manuscript. We hope that the revised version now fully meets the journal's standards for publication.

Comment # 2:

2. Throughout the paper, many claims made by the authors are unsupported by citations to relevant works (see below for several examples).

Response:

- We sincerely thank you for pointing out this important issue, and we apologize for the insufficient citation support in the previous version of the manuscript. Following your suggestion, we have carefully reviewed the entire paper and thoroughly examined every statement and reference to ensure that all claims are now properly supported by relevant and up-to-date literature.
- In addition to addressing each of your specific comments below, we have also revised and updated the reference list throughout the manuscript to improve citation accuracy and completeness.
- Once again, we appreciate your careful reading and constructive feedback, which have significantly improved the quality and credibility of our work.

Comment # 3:

2. The figures are helpful but there are not enough references to them in the text. I believe I counted exactly one in-text reference per figure, and often they seemed out of place. It would be helpful to refer the reader to the appropriate figure more often.

20. L289-290: The reference to Figure 2 here is confusing. You are in the middle of discussing DeepLab, but there is no mention of deeplab in figure 2.

Response:

- Thank you for reviewing our manuscript and providing valuable and constructive feedback. Your observation that the manuscript contained insufficient and sometimes awkwardly placed references to figures and tables is highly insightful.
- We fully agree with your assessment! In the revised version, we have systematically reviewed and refined the entire manuscript, substantially increasing the number of figure references and ensuring that each is seamlessly integrated with the corresponding textual discussion to better guide readers and strengthen our arguments.

Original Description

……

Ground-based sensing devices are highly diverse, and the data they acquire directly reflect the state of landslide masses. These datasets provide foundational inputs for deep learning models, enabling multi-dimensional analysis and interpretation of potential landslide conditions. For example, ground sensors (e.g., GNSS receivers and crack meters) can collect parameters like displacement and tilt angle at frequencies ranging from minutes to seconds, capturing transient, anomalous signals just prior to landslide events, thereby filling the temporal resolution gap in remote sensing (Jiang et al., 2022). By integrating time series data with SAR imagery, deep learning models can be trained to uncover correlation patterns between surface deformations and subsurface parameters. Instruments such as piezometers and soil pressure gauges can directly monitor key parameters like pore water pressure and soil stress on the sliding surface. By combining the obtained subsurface data with geomechanical equations, the position of the sliding surface or geotechnical strength parameters can be inverted.

……

Convolutional neural networks (CNNs) represent the fundamental architecture in image processing. A CNN primarily comprises convolutional layers, pooling layers, and fully connected layers, each performing predefined functions on its input data (Kattenborn et al., 2021; Liu et al., 2022a).

……

DenseNet is a further innovation of ResNet (Huang et al., 2017). Both of these neural networks are based on a similar idea, which is to establish a "shortcut" between different layers. However, the structure of DenseNet is simpler and more effective, with fewer parameters. In ResNet, each layer is only connected to the previous layer, while in DenseNet, each layer is directly connected to all previous layers, and each layer can obtain gradients from the loss function. This can optimize the information flow and gradients of the entire network, making it easier to train and performing better on small datasets. The structure of DenseNet can achieve better feature reuse and reduce the number of parameters. Moreover, the layers of DenseNet are narrower than those of other deep learning networks (Liu et al., 2021c), making it reduce redundancy by learning with fewer feature maps. This architecture is suitable for the extraction of multi-scale landslide features under complex terrains, even with limited landslide training samples.

……

In contrast, the U-Net architecture is relatively simpler and better suited for small targets and high-resolution imagery, such as landslide crack segmentation or fine annotation of high-resolution UAV images. DeepLab, on the other hand, is more effective for large-scale landslide area detection and multispectral remote sensing image classification (see Fig. 2).

……

Different from conventional statistical or physical models, deep learning models can automatically reveal dynamic change trends and periodic patterns in the data, providing more accurate information for landslide prediction and early warning. Recently, deep learning–based temporal models, such as recurrent neural networks (RNNs), long short-term memory networks (LSTMs), and gated recurrent units (GRUs), have become key tools for extracting nonlinear dependencies and temporal evolution patterns in landslide-related time series.

……

To overcome the vanishing gradient problem inherent in RNNs, LSTMs introduce memory cells and gating mechanisms that selectively retain relevant temporal information (Landi et al., 2021; Sherstinsky, 2020; Smagulova and James, 2019; Staudemeyer and Morris, 2019; Yu et al., 2019).

This capability allows them to model the cumulative and delayed responses of slopes to prolonged rainfall or reservoir water level fluctuations.

……

Deep generative models are the leading deep learning approach for synthetic data generation (Alam et al., 2018; Karras et al., 2020; Ma et al., 2024; Xu et al., 2015). They utilize deep neural networks to learn latent representations of data and optimize the learning process through specific objective functions. A key characteristic of deep generative models lies in their probabilistic nature—they not only classify or reconstruct data but also capture the underlying distribution of geospatial features, enabling the creation of new landslide samples that are statistically consistent with observed patterns. Commonly used deep generative models include generative adversarial networks (GANs), variational autoencoders (VAEs), and diffusion models.

……

GANs consist of a generator and a discriminator that compete in an adversarial process (Goodfellow et al., 2014). The generator synthesizes data resembling real samples, while the discriminator attempts to distinguish between generated and real data. Through iterative adversarial training, the generator learns to produce high-quality synthetic data that closely matches the distribution of real data (Gui et al., 2021; Saxena and Cao, 2021).

……

When computational resources and training time permit, diffusion models provide a powerful alternative for generating high-quality, diverse, and stable data (Ho et al., 2020; Croitoru et al., 2023; Yang et al., 2023a; Zhu et al., 2023a). These models learn the data distribution by gradually adding noise to real samples (forward diffusion) and then reconstructing clean data through a reverse denoising process. The resulting models can sample new, realistic data points that reflect complex terrain and geophysical variability.

……

As previously introduced, VAE a probabilistic extension of AEs. VAEs introduce stochastic latent variables characterized by mean and variance, allowing them to model data uncertainty (Kingma et al., 2013; Li et al., 2020; Park et al., 2018). During training, VAEs learn the latent distribution of normal samples and reconstruct inputs accordingly. When new observation data deviate significantly from the learned distribution, the reconstruction error increases accordingly, and this phenomenon can be used as an indicator of potential anomalies.

……

Transformer architectures, characterized by the self-attention mechanism, provide another promising avenue for landslide-related data fusion (Huang and Chen, 2023; Zhao et al., 2021a). Unlike CNNs or RNNs, which process spatial or temporal sequences sequentially, Transformers can jointly capture long-range dependencies across spatial and temporal dimensions, enabling unified processing of rainfall, InSAR time series, and topographic data (Esser et al., 2021; Lv et al., 2023).

……

In this section, we focus on the analysis of the second type of potential landslides. Based on triggering factors, landslides can be classified into four categories: rainfall-induced landslides, earthquake-induced landslides, human activity-induced landslides, and multi factor-induced landslides. For each category of landslide, we provide a brief outline of its characteristics, discuss the applications of deep learning to different types of landslides, and examine the selection of monitoring methods for each category.

……

Multi factor-induced landslides result from the synergistic interaction of multiple natural and anthropogenic factors (Hao et al., 2023). Their triggering mechanisms involve the dynamic spatiotemporal coupling of these factors, driving progressive destabilization of geomaterials through cumulative strength degradation. The formation of such landslides may involve various types of movements, including collapse, creep, and flow phenomena. They often exhibit characteristics such as complexity, nonlinearity, and suddenness. Therefore, their identification is markedly more complex compared to landslides triggered by singular factors.

……

The abstract features extracted by the models also lack correspondence to interpretable geological indicators. Even if the model can identify potential landslides through the texture patterns of remote sensing images, it cannot explain whether these patterns correspond to the actual geomechanical parameters.

……

In reality, the collection of landslide inventories faces many difficulties and it is hard to obtain them comprehensively and accurately. Thus, data scarcity is a common problem in the identification of potential landslide, especially in remote areas or regions with limited data accessibility. In such cases, deep learning models may suffer from overfitting or insufficient generalization ability due to a lack of samples (Kong et al., 2025; Lee et al., 2018). Although there are large-scale datasets such as the CAS landslide dataset, they are still insufficient compared with the data requirements of deep learning models (Xu et al., 2024).

……

Although deep learning models have achieved success in landslide identification, they also have certain problems of their own. The most critical challenge is interpretability (Li et al., 2025). This means that it is difficult to explain how these models achieve these results.

……

The formation of landslides involves the dynamic coupling of multiple factors such as geological structures, geotechnical mechanics, hydrological conditions, topography, meteorological factors, vegetation coverage, and human activities (Scheingross et al., 2020). Therefore, the triggering mechanisms encompass multiscale processes spanning microscopic interparticle friction to macroscopic slope instability, and transient dynamic responses to long-term temporal evolution (Yi et al., 2022).

The knowledge-data dually driven paradigm, operating through an iterative "theory-guided data assimilation and data informed theoretical refinement" mechanism, has advanced potential landslide identification from empirical reliance to scientifically quantifiable methodologies.

……

Revised Description

……

Ground-based sensing devices are highly diverse, and the data they acquire directly reflect the state of landslide masses. These datasets provide foundational inputs for deep learning models, enabling multi-dimensional analysis and interpretation of potential landslide conditions. For example, ground sensors (e.g., GNSS receivers and crack meters) can collect parameters like displacement and tilt angle at frequencies ranging from minutes to seconds (Jiang et al., 2022), capturing transient, anomalous signals just prior to landslide events, thereby filling the temporal resolution gap in remote sensing (see Figure 1). By integrating time series data with SAR imagery, deep learning models can be trained to uncover correlation patterns between surface deformations and subsurface parameters. Instruments such as piezometers and soil pressure gauges can directly monitor key parameters like pore water pressure and soil stress on the sliding surface. By combining the obtained subsurface data with geomechanical equations, the position of the sliding surface or geotechnical strength parameters can be inverted.

……

Convolutional neural networks (CNNs) represent the fundamental architecture in image processing. A CNN primarily comprises convolutional layers, pooling layers, and fully connected layers, each performing predefined functions on its input data, as illustrated in Fig. 2 (Kattenborn et al., 2021; Liu et al., 2022a).

……

DenseNet is a further innovation of ResNet (Huang et al., 2017). Both of these neural networks are based on a similar idea, which is to establish a "shortcut" between different layers. However, the structure of DenseNet is simpler and more effective, with fewer parameters. The structural differences between ResNet and DenseNet are illustrated in Fig. 2. In ResNet, each layer is only connected to the previous layer, while in DenseNet, each layer is directly connected to all previous layers, and each layer can obtain gradients from the loss function. This can optimize the information flow and gradients of the entire network, making it easier to train and performing better on small datasets. The structure of DenseNet can achieve better feature reuse and reduce the number of parameters. Moreover, the layers of DenseNet are narrower than those of other deep learning networks (Liu et al., 2021c), making it reduce redundancy by learning with fewer feature maps. This architecture is suitable for the extraction of multi-scale landslide features under complex terrains, even with limited landslide training samples.

……

In contrast, the U-Net architecture is relatively simpler and better suited for small targets and high-resolution imagery, such as landslide crack segmentation or fine annotation of high-resolution UAV images. DeepLab, on the other hand, is more effective for large-scale landslide area detection and multispectral remote sensing image classification.

……

Different from conventional statistical or physical models, deep learning models can automatically reveal dynamic change trends and periodic patterns in the data, providing more accurate information for landslide prediction and early warning. Recently, deep learning–based temporal models, such as recurrent neural networks (RNNs), long short-term memory networks (LSTMs), and gated recurrent units (GRUs), have become key tools for extracting nonlinear dependencies and temporal evolution patterns in landslide-related time series. The structural characteristics and differences among these models are illustrated in Fig. 3.

……

To overcome the vanishing gradient problem inherent in RNNs, LSTMs introduce memory cells and gating mechanisms that selectively retain relevant temporal information (Landi et al., 2021; Sherstinsky, 2020; Smagulova and James, 2019; Staudemeyer and Morris, 2019; Yu et al., 2019). As shown in Fig. 3, LSTM networks extend the basic RNN structure by incorporating gating units that control information flow, enabling them to better capture cumulative and delayed slope responses to environmental triggers. This capability allows them to model the cumulative and delayed responses of slopes to prolonged rainfall or reservoir water level fluctuations.

……

Deep generative models are the leading deep learning approach for synthetic data generation (Alam et al., 2018; Karras et al., 2020; Ma et al., 2024; Xu et al., 2015). They utilize deep neural networks to learn latent representations of data and optimize the learning process through specific objective functions. A key characteristic of deep generative models lies in their probabilistic nature—they not only classify or reconstruct data but also capture the underlying distribution of geospatial features, enabling the creation of new landslide samples that are statistically consistent with observed patterns. Commonly used deep generative models include generative adversarial networks (GANs), variational autoencoders (VAEs), and diffusion models (see Figure 4).

……

GANs consist of a generator and a discriminator that compete in an adversarial process (Goodfellow et al., 2014). The generator synthesizes data resembling real samples, while the discriminator attempts to distinguish between generated and real data. The workflow of adversarial training for GAN-based data generation is schematically depicted in Fig. 4. Through iterative adversarial training, the generator learns to produce high-quality synthetic data that closely matches the distribution of real data (Gui et al., 2021; Saxena and Cao, 2021).

……

When computational resources and training time permit, diffusion models provide a powerful alternative for generating high-quality, diverse, and stable data (Ho et al., 2020; Croitoru et al., 2023; Yang et al., 2023a; Zhu et al., 2023a). These models learn the data distribution by gradually adding noise to real samples (forward diffusion) and then reconstructing clean data through a reverse denoising process (see Figure 4). The resulting models can sample new, realistic data points that reflect complex terrain and geophysical variability.

……

As previously introduced, VAE a probabilistic extension of AEs (see Figure 4). VAEs introduce stochastic latent variables characterized by mean and variance, allowing them to model data uncertainty (Kingma et al., 2013; Li et al., 2020; Park et al., 2018). During training, VAEs learn the latent distribution of normal samples and reconstruct inputs accordingly. When new observation data deviate significantly from the learned distribution, the reconstruction error increases accordingly, and this phenomenon can be used as an indicator of potential anomalies.

……

Transformer architectures, characterized by the self-attention mechanism (see Figure 5), provide another promising avenue for landslide-related data fusion (Huang and Chen, 2023; Zhao et al., 2021a). Unlike CNNs or RNNs, which process spatial or temporal sequences sequentially, Transformers can jointly capture long-range dependencies across spatial and temporal dimensions, enabling unified processing of rainfall, InSAR time series, and topographic data (Esser et al., 2021; Lv et al., 2023).

……

In this section, we focus on the analysis of the second type of potential landslides. Based on triggering factors, landslides can be classified into four categories: rainfall-induced landslides, earthquake-induced landslides, human activity-induced landslides, and multi factor-induced landslides (see Figure 6). For each category of landslide, we provide a brief outline of its characteristics, discuss the applications of deep learning to different types of landslides, and examine the selection of monitoring methods for each category.

……

Multi factor-induced landslides result from the synergistic interaction of multiple natural and anthropogenic factors (Hao et al., 2023). Their triggering mechanisms involve the dynamic spatiotemporal coupling of these factors, driving progressive destabilization of geomaterials through cumulative strength degradation (see Figure 6). The formation of such landslides may involve various types of movements, including collapse, creep, and flow phenomena. They often exhibit characteristics such as complexity, nonlinearity, and suddenness. Therefore, their identification is markedly more complex compared to landslides triggered by singular factors.

……

The abstract features extracted by the models also lack correspondence to interpretable geological indicators (see Figure 8). Even if the model can identify potential landslides through the texture patterns of remote sensing images, it cannot explain whether these patterns correspond to the actual geomechanical parameters.

……

In reality, the collection of landslide inventories faces many difficulties and it is hard to obtain them comprehensively and accurately. Thus, data scarcity is a common problem in the identification of potential landslide, especially in remote areas or regions with limited data accessibility (as illustrated in Fig. 8). In such cases, deep learning models may suffer from overfitting or insufficient generalization ability due to a lack of samples (Kong et al., 2025; Lee et al., 2018). Although there are large-scale datasets such as the CAS landslide dataset, they are still insufficient compared with the data requirements of deep learning models (Xu et al., 2024).

……

Although deep learning models have achieved remarkable success in landslide identification, they also face inherent limitations (Li et al., 2025), particularly regarding interpretability, robustness, and generalization (see Figure 8). The most critical challenge is interpretability, meaning it is difficult to explain how these models achieve their results.

……

The formation of landslides involves the dynamic coupling of multiple factors such as geological structures, geotechnical mechanics, hydrological conditions, topography, meteorological factors, vegetation coverage, and human activities (Scheingross et al., 2020; Yi et al., 2022). Therefore, the triggering mechanisms encompass multiscale processes spanning microscopic interparticle friction to macroscopic slope instability, and transient dynamic responses to long-term temporal evolution (see Figure 8).

……

The knowledge-data dually driven paradigm, operating through an iterative "theory-guided data assimilation and data informed theoretical refinement" mechanism, has advanced potential landslide identification from empirical reliance to scientifically quantifiable methodologies. The overall workflow of this knowledge-data dually driven paradigm for potential landslide identification is conceptually summarized in Figure 9.

……

Comment #4:

3. I felt the paper provided good background on landslides, and good background on deep learning, but was missing emphasis on the intersection: the application of deep learning to landslides. I would have expected this to appear in Section 3 or 4. However, section 3 discusses deep learning methods (without much specific discussion of how previous works have utilized them for landslides), and section 4 is almost entirely about the actual mechanisms of landslides, rather than how deep learning can be beneficial here. Section 4 seems almost mis-titled in this regard. To be precise, discussions similar to lines 632-638 are what I would have expected to see more of: specific examples of specific methods applied to specific problems in landslide applications.

Response:

- Thank you for accurately identifying the core weakness of our manuscript and for providing a clear direction for improvement. We fully agree with your assessment that the original version did not effectively bridge the two domains.
- To thoroughly address this issue, we have undertaken substantial revisions:
- (1) We have completely reorganized and rewritten Section 3. Instead of adopting a general, tutorial-style discussion of deep learning models, the revised section is structured around specific landslide-related tasks. In each subsection, we explicitly describe the specific deep learning architectures employed to address these tasks, accompanied by numerous relevant studies as concrete examples. We further elaborate on how these models were applied, as well as their respective advantages and limitations. (Please see the new **Section 3** for details).
- (2) We have also substantially expanded Section 4 by incorporating numerous real-world applications. For each type of landslide discussed, we now integrate specific research cases to analyze which deep learning models are most suitable for that particular type. (Please see **Comment #27**).

- Original Description in Section 3

3.1 Models for Image Analysis and Processing in Potential Landslide Identification

Image data plays a critical role in potential landslide identification, especially through remote sensing, satellite, and UAV imagery. These images enable the acquisition of large-scale terrain data, encompassing complex geographical features, vegetation coverage, and ground fissures, which often serve as potential precursors to landslide occurrences. The adoption of deep learning has facilitated a shift from conventional manual visual interpretation to automated high-precision segmentation.

Convolutional neural networks (CNNs) represent the fundamental architecture in image processing. A CNN primarily comprises convolutional layers, pooling layers, and fully connected layers, each performing predefined functions on its input data (Kattenborn et al., 2021; Liu et al., 2022a).

The convolutional layer, as the core component of CNNs, contains multiple kernels that progressively extract more detailed feature representations (Hussain et al., 2019; Shi et al., 2020; Yao et al., 2021). Meanwhile, the shared-weight strategy inherent in convolutional layers allows for network training with fewer parameters than fully connected architectures. Convolutional kernels of different sizes facilitate multi-scale feature extraction. Small kernels focus on fine details, such as small cracks and the texture of localized soil loosening, while large kernels emphasize capturing overall shapes, such as the general outline of landslides and the macroscopic morphology of mountain bodies. Pooling layers, typically positioned after convolutional layers, serve to reduce the size of feature representations and enhancing the model's resistance to overfitting when handling diverse data. Common pooling methods include max pooling and average pooling, which enhance robustness to minor transformations such as translation and rotation, ensuring a degree of invariance in the features extracted by CNNs. Pooling operations downsample the convolved feature maps, reducing computational complexity while reinforcing feature robustness. Through the hierarchical stacking of multiple convolutional and pooling layers, CNNs incrementally extract more abstract and semantically rich features (Mao et al., 2024). The final fully connected layer flattens the pooled feature maps and performs classification, outputting results that distinguish potential landslide areas from non-landslide areas or enable further analysis of landslide types (Wu et al., 2024).

The layers of a CNN can be combined in various ways, forming distinct CNN architectures. These architectures are primarily determined by task requirements, which may include image classification, multi-class segmentation, or object localization within a scene.

Conventional CNNs typically consist of multiple stacked convolutional layers, pooling layers, and fully connected layers. However, increasing network depth introduces challenges such as vanishing gradients and degradation arise, resulting in model performance deterioration.

ResNet addresses these limitations by integrating residual blocks into the foundational CNN framework (Qi et al., 2020; Yang et al., 2022). These residual blocks utilize shortcut connections that preserve original feature information. This framework facilitates the construction of ultra-deep networks capable of extracting high-level semantic features for landslide detection, thereby enhancing adaptability to complex terrain classification tasks (Ullo et al., 2021). Models with higher parameter counts generally exhibit greater representational capacity but are prone to overfitting, while demanding higher computational resources and temporal costs for both training and inference. For instance, ResNet-152 contains orders of magnitude more parameters than ResNet-50, yet the latter is often preferable in computationally constrained environments due to its balanced efficiency and performance.

DenseNet is a further innovation of ResNet (Huang et al., 2017). Both of these neural networks are based on a similar idea, which is to establish a "shortcut" between different layers. However, the structure of DenseNet is simpler and more effective, with fewer parameters. In ResNet, each layer is only connected to the previous layer, while in DenseNet, each layer is directly connected to all previous layers, and each layer can obtain gradients from the loss function. This can optimize the information flow and gradients of the entire network, making it easier to train and performing better on small datasets. The structure of DenseNet can achieve better feature reuse and reduce the number of parameters. Moreover, the layers of DenseNet are narrower than those of other deep learning networks (Liu et al., 2021c), making it reduce redundancy by learning with fewer feature maps. This architecture is suitable for the extraction of multi-scale landslide features under complex terrains, even with limited landslide training samples.

With the rapid expansion of deep learning methods based on CNNs, semantic segmentation models have increasingly become the standard in landslide detection. Numerous advanced semantic segmentation networks have been proposed and validated for automatic landslide detection, significantly enhancing the efficiency and accuracy of large-scale detection. U-Net is a typical example (Ronneberger et al., 2015), which features a U-shaped architecture. U-Net employs an encoder-decoder structure, where the encoder is similar to conventional CNNs, progressively reducing image resolution and extracting features through convolution and pooling operations; the decoder then restores the image resolution through transposed convolution or upsampling operations (Dong et al., 2022; Nava et al., 2022). Skip connections bridge low-level detail features with deep semantic features, thereby refining segmentation precision.

When dealing with complex features in landslide-prone areas, DeepLab is a more suitable choice (Sandric et al., 2024). Built upon deep convolutional neural networks, DeepLab employs dilated convolutions to expand the receptive field and integrates an atrous spatial pyramid pooling (ASPP) module to capture multi-scale contextual information.

In contrast, the U-Net architecture is relatively simpler and better suited for small targets and high-resolution imagery, such as landslide crack segmentation or fine annotation of high-resolution UAV images. DeepLab, on the other hand, is more effective for large-scale landslide area detection and multispectral remote sensing image classification (see Fig. 2).

After achieving semantic segmentation to obtain the accurate extent of a landslide and the classification of ground objects, change detection is employed to monitor the changes in the landslide area over time. By comparing the segmentation results of multiple temporal phases or directly analyzing the feature differences, the dynamic evolution of potential hazards can be quantifie (Amankwah et al., 2022).

Wang (2023) demonstrates that 3D CNNs can directly process these 3D tensors. These models capture spatial and temporal features using convolutional kernels while transforming multi-temporal image sequences into change hotspot maps or temporal variation curves as output.

Some studies even have integrated attention mechanisms into conventional CNN architectures to enhance the analysis of multi-temporal remote sensing imagery, thereby enabling the identification of landslide hazard evolution over time.

3.2 Models for Time Series Analysis in Potential Landslide Identification

The occurrence of a landslide is a gradual accumulation process, usually influenced by a variety of factors. We refer to data that reflect the changing states of a landslide body over time as time series data. Time series data analysis aims to excavate the information hidden in the time series data to help identify potential landslides. Different from conventional time series data analysis methods, using deep learning models an automatically reveal the dynamic change trends and periodic patterns in the data, providing more accurate information for landslide prediction.

Recurrent neural networks (RNNs) are a class of deep learning models specialized in processing sequential data, capable of capturing temporal dependencies within input sequences (Ngo et al., 2021; Zaremba et al., 2014). Unlike conventional feedforward neural networks, in an RNN, each neuron not only receives the current input but also the output of the previous time step as additional input. This structure endows the RNN with a memory mechanism.

The architecture contains three primary components working in coordination:(1) The input layer means that one data point is input at each time step. (2) The hidden layer contains recurrent connections, which enable the information from the previous time step to be passed to the current time step, and the output serves as the input for the next time step simultaneously. (3) The output layer generates the output under the control of the state of the hidden layer (Cho et al., 2014; Zhao et al., 2021b).

During the training process, the RNN will process the data at each time step in sequence, continuously updating the hidden state. By combining the input of the current time step with the hidden state of the previous moment for calculation to gain an understanding of the data at the current moment, this structure enables the RNN to capture the temporal evolution patterns of landslide-related factors.

Due to conventional RNNs struggle to model long-term dependencies and limit their applicability to short-term temporal sequences, long short-term memory networks (LSTM) were developed (Wang et al., 2023b).

LSTM is an enhancement of RNNs, primarily processing long sequence data (Hochreiter and Schmidhuber, 1997). Compared to standard RNNs, the hidden layer architecture of LSTM is much more complex. By incorporating memory cells and gating mechanisms, LSTM selectively propagates critical information across multiple time steps, thereby effectively capturing long-range temporal dependencies (Landi et al., 2021; Yu et al., 2019).

The basic unit of an LSTM consists of three primary gates: (1) the input gate, which determines what new information should be added to the cell state; (2) the forget gate, which decides what old information should be discarded; and (3) the output gate, which selects the information to be output from the cell state as the hidden state at the current time step (Sherstinsky, 2020; Smagulova and James, 2019; Staudemeyer and Morris, 2019). The output hidden state, after a nonlinear transformation, can be used for prediction or as the input for the next time step (Yang et al., 2019).

This structure allows the LSTM to retain key information over long sequences while selectively forgetting irrelevant information according to the requirements. Through learning from historical data, the LSTM can predict the likelihood of landslides occurring, as well as the possible scale and impact range of landslides under different future conditions.

Due to the ability to self-update weights and significantly improve network accuracy, LSTMs can also be used as a complex nonlinear component in the construction of larger deep neural networks. The model does not require separating trend and periodic components from the original deformation data, yet it can compensate for deformation trend predictions caused by unexpected interruptions in monitoring data. These properties make LSTMs particularly suited for high-accuracy research and analytical scenarios requiring large-scale datasets (Gidon et al., 2023; Xu and Niu, 2018).

Gated recurrent unit (GRU) is a simplified version of LSTM(Chung et al., 2014; Zhang et al., 2022b), which has fewer parameters. Due to their higher computational efficiency, GRU has potential advantages in real-time data processing scenarios in landslide monitoring.

GRU mainly consists of the update gate and reset gate. The update gate is used to control how much of the previous information should be preserved at the current time step, while the reset gate is used to determine whether to ignore the hidden state of the previous time step, enabling the model to adaptively learn information across different temporal scales. This dual-gate mechanism enables adaptive learning of multi-scale temporal patterns.

Compared with the LSTM, the GRU has fewer parameters and higher computational efficiency, giving it an advantage in some landslide monitoring scenarios where real-time performance is critical.

GRU is capable of effectively handling time series data with long-term dependencies, making it suitable for long-term prediction of landslide hazards. Moreover, by learning temporal patterns in historical data, GRU can identify critical conditions for landslide occurrence in advance. GRU particularly well-suited for applications involving real-time analysis of on-site monitoring data, where rapid detection of imminent landslide risks is essential and data volume is relatively limited.

Transformer was originally designed to handle sequential data in natural language processing, which was first introduced by Vaswani in 2017 (Vaswani et al., 2017). Unlike conventional recurrent and convolutional structures, the Transformer employs employs a self-attention mechanism to directly model the entire sequence.

Since the Transformer has the ability to adaptively learn latent features and patterns within the data, when it comes to processing landslide time series data, it can automatically tweak the model parameters to accommodate diverse landslide scenarios and temporal data variability (Wang et al., 2024a; Zerveas et al., 2021).

Transformer also can analyze positional relationships across the entire sequence, better capturing complex dependencies in long sequences, making it especially suitable for handling large-scale, long-term sequential datasets.

In contrast, RNN-based models exhibit a relatively simple architecture (Li et al., 2021a; Wang et al., 2020b). Their mechanisms are conceptually intuitive, making them more interpretable (see Fig. 3). On the other hand, Transformers are more complex in structure with numerous parameters, necessitating substantial computational resources during early training to process large-scale data, while being susceptible to overfitting on small datasets. Understanding how the model extracts features and makes decisions is not straightforward from large amounts of landslide data, posing challenges for its interpretability and practical deployment.

3.3 Models for Data Generation in Potential Landslide Identification

Data generation refers to modeling the underlying data distribution of data to generate entirely new samples independent of the original dataset (Kingma et al., 2014; Moreno-Barea et al., 2020; Shorten and Khoshgoftaar, 2019), thereby enriching the dataset. In potential landslide identification, data generation mitigates challenges of data scarcity and imbalanced class distributions, thereby enhancing the generalization capability of predictive models.

Deep generative models are the leading deep learning approach for synthetic data generation

(Alam et al., 2018; Karras et al., 2020; Ma et al., 2024; Xu et al., 2015). They operate on principles similar to those of deep learning, utilizing deep neural networks to learn data representations and optimizing the learning process through objective functions.

A fundamental characteristic of deep generative models lies in their probabilistic nature. These models learn an approximate probability distribution from observed samples and subsequently generate novel samples that maintain statistical consistency with the original dataset. Unlike conventional discriminative models, generative models not only classify data but also learn the underlying distribution and generate new data points. Commonly used deep generative models include generative adversarial networks (GANs), variational autoencoders (VAEs, a variant of autoencoders), and diffusion models.

GAN is a suitable choice to generate highly realistic and diverse new images (Goodfellow et al., 2014; Tran et al., 2021). Instead of explicitly modeling data distributions, GANs implicitly learn distributions through adversarial training between generator and discriminator networks.

During data generation, the generator network in a GAN synthesizes images or data resembling real samples by processing input noise vectors (Gui et al., 2021; Saxena and Cao, 2021). The discriminator, on the other hand, is used to distinguish between the generated data and the real data.These two components are continuously optimized through adversarial training. Eventually, the generator is able to produce high-quality synthetic data, which is highly similar to the real data in terms of features.

With this adversarial structure (Al-Najjar et al., 2021), GANs can generate high-quality data that closely matches the distribution of real data in an unsupervised learning context, making them well-suited for high-resolution image synthesis.

With the proposal and development of GANs, researchers have introduced various enhanced structures that are more effectively applied to potential landslide identification. For example, the conditional GAN (CGAN) (Kim and Lee, 2020; Loey et al., 2020), Pix2Pix (Qu et al., 2019), and Wasserstein GAN (WGAN) (Wang et al., 2019).

In the case of GANs, although the generated high-quality images may visually resemble real potential landslide regions, mode collapse can lead to a lack of diversity in the generated data, failing to cover all possible types of hazards (Fang et al., 2020). If certain types of potential landslides are underrepresented in the training dataset, GANs may struggle to generate those types effectively, thereby limiting the effectiveness of data augmentation. Given that the inherently unstable training process of the GANs may require more hyperparameter tuning and computational resources, this model will pose additional challenges in scenarios with limited data availability (Al-Najjar and Pradhan, 2021; Feng et al., 2024).

As a variant of the autoencoders (AEs), the variational autoencoder (VAE) introduces the idea of probabilistic generation (Kingma et al., 2013). VAE constrains the latent space through variational inference, thus enabling the generation, reconstruction, and transformation of sample data.

Compared to GANs, the samples generated by the VAE may have better diversity (Cai et al., 2024; Islam et al., 2021; Oliveira et al., 2022), because the structured constraints of its latent space are helpful for generating samples with continuous changes. This is beneficial for simulating potential landslides under different geological conditions.

The encoder of the VAE maps the input data to a low-dimensional latent space, where each vector represents the underlying features of the input. The decoder then reconstructs the original data based on the vectors in the latent space. Different from conventional AEs, the output of the VAE encoder includes two parameters: the mean value and the standard deviation. These two parameters define the probability distribution in the latent space, which is usually assumed to be a Gaussian distribution. The decoder samples a latent variable from this probability distribution and reconstructs it into output data, thus generating data with inherent randomness and diversity. Therefore, the VAE can extract latent features from landslide data and generate new landslide data based on these features.

By learning from extensive landslide datasets, VAEs capture critical geomorphological features and patterns, enabling the generation of novel samples that preserve these characteristics. This capability enables innovative applications in potential landslide analysis. This is crucial for exploring landslide scenarios under different feature combinations and identifying potential landslide patterns. Compared to GANs, VAEs exhibit superior sample diversity and training stability though the generated samples often lack the fine-grained details produced by GANs, particularly in high-resolution geospatial contexts. Moreover, VAEs may still face challenges in handling highly imbalanced data, as the generated samples tend to favor majority classes, which can limit its effectiveness in augmenting minority class data.

When computational resources and time are sufficient, and high-quality data generation with exceptional diversity is prioritized, diffusion models are the recommended choice (Croitoru et al., 2023; Yang et al., 2023a; Zhu et al., 2023a).

Diffusion models fundamentally learn the distribution of data. During training, the model applies a forward diffusion process that gradually adds noise to the original data until it approximates a Gaussian distribution. Then, in the reverse diffusion process, the model learns to iteratively refine its reconstruction of the original data distribution from the noisy data. After being fully trained, the model is able to capture the latent distribution patterns of the data, and thus can sample based on the learned distribution to generate new data (Ho et al., 2022). That is to say, by grasping the inherent laws and features of the data, the model has the ability to generate data that conforms to the distribution of the data.

Denoising diffusion probabilistic model (DDPM) is a classic implementation of the diffusion models, which lays the probabilistic framework for the diffusion models (Choi et al., 2021; Ho et al., 2020; Jing et al., 2023; Perera et al., 2023). The generation quality is optimized through variational inference and noise scheduling. Denoising diffusion implicit model (DDIM) has made improvements on the basis of DDPM (Song et al., 2020). It uses non-Markov chain reparameterization and deterministic sampling, and greatly improves the efficiency with almost no loss of quality.

Notably, DDIMs utilize the same training framework as DDPMs. If certain parameters of DDIMs are assigned particular values, its generation process becomes equivalent to DDPMs. Thus, DDIMs function as an accelerated sampling variant of DDPMs. The critical distinction lies in their sampling mechanisms. DDPMs employ stochastic and Markovian sampling, whereas DDIMs enhance efficiency through non-Markovian deterministic sampling, though this comes at the expense of reduced sample diversity.

Although diffusion models demonstrate strong capabilities in generating high-quality images and handling noise, they generate superior-quality data and ensure greater training stability compared to GANs and VAEs. However, diffusion models have not yet been widely applied directly to the identification of potential landslides and remain in the exploratory stage (see Fig. 4). We believe that as generative models advance in the field of geospatial remote sensing, they hold vast potential for application and could play a pivotal role in future landslide risk analysis and monitoring systems.

3.4 Models for Data Cleaning in Potential Landslide Identification

In potential landslide identification, data cleaning, particularly anomaly detection, is a critical issue (Deijns et al., 2020; Jiang et al., 2020). It can distinguish between normal fluctuations and true anomalies, identifying early signs such as subtle changes in the mountain’s state or abnormal trends in surface displacement, thus enabling more accurate landslide hazard assessment. With the rapid development of deep learning, the applications in data cleaning have become increasingly widespread, enabling models to automatically learn latent data patterns and identify potential anomalies.

AEs and their variational counterparts are highly effective in unsupervised data cleaning. These models autonomously learn normal geomechanical patterns from data and flag deviations, achieving effective hazard identification even when labeled anomaly samples are scarce.

The AE is a typical unsupervised learning model consisting of an encoder and a decoder The encoder compresses the input data into low-dimensional features, and then the decoder reconstructs the input. During the training process, the autoencoder learns the intrinsic features and patterns of normal landslide data, so that for normal data, the reconstruction error is small. When abnormal landslide data is input, due to the difference between its features and the distribution of normal data, the reconstruction error will be large.

When performing anomaly detection, a suitable reconstruction error threshold is set. When the reconstruction error of the test data exceeds this threshold, it can be determined as abnormal data. In the anomaly detection of landslide displacement data monitored by sensors, if the error of the displacement data after being reconstructed by AEs during a certain period is significantly higher than the normal level, it may indicate that there is an abnormal situation of potential landslides during this period.

As previously introduced, VAE is an extension of AE. Compared to conventional autoencoders, VAE introduces randomness into the latent space, making it more effective in handling data uncertainty (Li et al., 2020; Park et al., 2018).

During training, VAEs learn the latent distribution of the data and can generate new samples resembling the training set. When input samples deviate significantly from this learned distribution, the VAE fails to reconstruct them accurately, thereby flagging anomalies through elevated reconstruction errors. For landslide monitoring, if a VAE is trained on imagery of stable slopes, it internalizes stable terrain features. When an image significantly differs from the stable region, the model will produce a high reconstruction error, indicating the presence of anomalous data.

In contrast, AEs are well-suited for univariate anomaly detection, particularly for landslide precursor detection, while VAEs capture latent space distributions and are more effective for multivariate anomaly detection.

GANs can also be utilized in data cleaning (Kang et al., 2024; Xia et al., 2022). In data cleaning, the discriminator is trained to distinguish between generated data and real data. When new test data is input, if the discriminator struggles to determine whether it is real or generated data, the test data may significantly deviate from the distribution of normal data, indicating a potential anomaly. In landslide monitoring, data may be influenced by various factors, GANs demonstrate robustness by filtering out such interference, thereby enhancing data cleaning accuracy (Radoi, 2022).

AnoGAN extends conventional GANs by directly incorporating data cleaning as one of its primary objectives (Lin et al., 2023; Thomine et al., 2023). It introduces an additional encoder during training, which maps input data to the latent space. The difference between this latent vector and the latent vector of normal samples generated by the generator serves as the basis for data cleaning.

RNNs, LSTMs, and GRUs are also effective for identifying anomalous patterns in sequential data (Zhang et al., 2022a). In potential landslide identification, these models process time series inputs to learn normal temporal dynamics and trends. When new data deviates significantly from the normal patterns learned by the model, such deviations can be flagged as anomalies. However, these models are primarily used for time series data, performing data cleaning by predicting future values of the sequence. For instance, if displacement measurements exhibit abrupt deviations while rainfall remains within historical norms, the model detects such discrepancies by comparing observed values with predictions based on learned temporal dependencies.

3.5 Models for Data Fusion in Potential Landslide Identification

In practical applications, the identification of potential landslide hazards is a complex task that influences by multiple factors (Zhang et al., 2018). These factors are often reflected through different data sources. We can roughly divide heterogeneous data into four categories: image data, time series data, structured data, and textual data. Data fusion is essential for the accurate identification of potential landslides. In order to better identify potential landslides, data fusion is essential.

Since the features, scales, and resolutions of heterogeneous data are all different, currently, the powerful feature learning ability of deep learning models is often utilized to automatically capture the nonlinear relationships and high-order interaction information among these heterogeneous data.

Due to the complex non-Euclidean structural characteristics of the geological environment, topographic data and their spatial relationships related to landslide hazards, conventional methods such as CNNs have difficulty in handling these relationships. As a neural network architecture for processing graph-structured data, graph neural networks (GNNs) can effectively model such spatial relationships (Ying et al., 2018; Zeng et al., 2022). They can treat the nodes in the geographical space (such as different geographical location points) and their connection relationships (such as the distance between adjacent nodes, terrain undulations, etc.) as the structure of a graph for processing.

When dealing with heterogeneous data, GNNs support feature interaction between different types of nodes through the message passing mechanism, thereby eliminating redundancy and mutual exclusivity among data sources and enabling dynamic fusion of multi-modal features (Zhang et al., 2024d; Zhao et al., 2024b). By passing and aggregating information across nodes, GNNs can also conduct a detailed analysis of various heterogeneous data in local areas. This capability allows GNNs to capture subtle geological structural changes and detect localized anomalies inmonitoring data, providing advantages for analyzing local features and early signs of potential landslide movements.

By learning a large amount of landslide potential cases, GNNs can discover the general patterns and rules of landslides, thus having good generalization ability. When facing new and unseen regions or data, GNNs can predict and assess the potential landslides in those regions based on the knowledge they have already learned.

According to the differences in message passing and aggregation methods, GNNs have derived various variants. For example, graph convolutional network (GCN) is generated by generalizing the convolutional operation to graph-structured data (Sharma et al., 2022; Wang et al., 2020a), and graph attention network (GAT) dynamically weights the importance of neighboring nodes by introducing the attention mechanism (Yuan et al., 2022; Zhou and Li, 2021). The emergence of these new architectures makes GNN variants more targeted than conventional GNNs and suitable for modeling heterogeneous relationships. Currently, they are often used for weighted analysis of the impacts of different geographical factors on landslides.

Transformer is also composed of stacked encoders and decoders (see Fig. 5). However, unlike other architectures, the Transformer architecture introduces the self-attention mechanism (Zhao et al., 2021a), which is a crucial innovation. This enables the Transformer to automatically calculate a weight vector for each position in the input sequence based on the relationship between this position and other positions, so as to represent the importance of this position in the entire sequence. Such a weight vector can be regarded as the "attention distribution" of each position in the input sequence, that is, the model determines which positions in the sequence to focus on. By considering all positions in the input sequence simultaneously, Transformer is able to calculate the correlations between each position and other positions in the sequence in parallel (Esser et al., 2021; Huang and Chen, 2023; Zerveas et al., 2021), rather than processing them step by step like CNNs or RNNs.

Transformer can also convert multimodal dFor different types of data, it transforms them into vector representations via different embedding layers.ata into a unified vector representation through different embedding layers. Subsequently, through the use of the self-attention mechanism and multilayer neural networks, these vectors are fused and feature representations are extracted, enabling the model to process and integrate data from various modalities within the same model framework (Lv et al., 2023; Tang et al., 2022).

Revised Description in Section 3

3.1 Models for Image Analysis and Processing in Potential Landslide Identification

Image data plays a critical role in potential landslide identification, especially through remote sensing, satellite, and UAV imagery. These images enable the acquisition of large-scale terrain data, encompassing complex geographical features, vegetation coverage, and ground fissures, which often serve as potential precursors to landslide occurrences. The adoption of deep learning has facilitated a shift from conventional manual visual interpretation to automated high-precision segmentation.

Convolutional neural networks (CNNs) represent the fundamental architecture in image processing (LeCun et al., 1998). A CNN primarily comprises convolutional layers, pooling layers, and fully connected layers, each performing predefined functions on its input data (Kattenborn et al., 2021; Liu et al., 2022a).

Convolutional layers, the core of CNNs, use kernels of various sizes to extract multi-scale features from geospatial imagery, which is crucial for landslide identification (Hussain et al., 2019; Shi et al., 2020; Yao et al., 2021). Small kernels are effective in detecting fine-grained precursors such as ground fissures and localized soil texture changes. For instance, Hamaguchi et al. (2018) proposed a Local Feature Extraction (LFE) module to enhance the capability of CNNs in identifying small object instances in remote sensing imagery. Wang et al. (2024) demonstrated the exceptional capability of convolutional layers in extracting extremely small and subtle features by identifying cracks as narrow as 0.05 m width using a U-Net–based model. In contrast, larger kernels help in recognizing the overall morphology and boundaries of landslide bodies. From the perspective of general visual tasks, Ding et al. (2022) demonstrated that larger convolution kernels substantially improve the shape bias of CNNs, facilitating the recognition of large-scale structures and overall morphological patterns compared with using small kernels alone. Li et al. (2025) employed multiple large convolution kernels (kernel sizes = 5, 7, and 9) within the DLFFSKA module to fuse multi-scale features, thereby enhancing the global perception of landslide boundaries and morphology as well as the capture of contextual background information.

Pooling layers downsample feature maps, improving computational efficiency and model robustness. In landslide mapping, this translation invariance is particularly beneficial, as it allows the model to consistently identify landslide features regardless of their slight positional variations across different image patches (Mao et al., 2024).

The final fully connected layer flattens the pooled feature maps and performs classification, outputting results that distinguish potential landslide areas from non-landslide areas or enable further analysis of landslide types (Wu et al., 2024).

The layers of a CNN can be combined in various ways, forming distinct CNN architectures. These architectures are primarily determined by task requirements, which may include image classification, multi-class segmentation, or object localization within a scene.

Conventional CNNs typically consist of multiple stacked convolutional layers, pooling layers, and fully connected layers. However, increasing network depth introduces challenges such as vanishing gradients and degradation arise, resulting in model performance deterioration.

ResNet mitigates the vanishing gradient problem in very deep networks through residual connections (He et al., 2016; Qi et al., 2020; Yang et al., 2022)). This architectural advancement has been successfully applied to landslide detection in complex terrains, such as the work by Ullo et al. (2021), who demonstrated that a ResNet-based classifier could achieve high accuracy in distinguishing landslide scars from surrounding vegetation and bare soil in satellite imagery by effectively learning hierarchical features.

Models with higher parameter counts generally exhibit greater representational capacity but are prone to overfitting, while demanding higher computational resources and temporal costs for both training and inference (Ebrahimi et al., 2021). For instance, He et al. (2016) introduced ResNet-152 and other deep residual network architectures, demonstrating that deeper structure achieve superior performance compared with shallower counterparts. Hasanah et al. (2023) explicitly highlighted the differences in layer depth and parameter count among various ResNet versions (ResNet-50, 101, and

152), noting that the increased number of parameters in deeper networks inevitably leads to longer training times.

DenseNet is a further innovation of ResNet (Huang et al., 2017). Both of these neural networks are based on a similar idea, which is to establish a "shortcut" between different layers. However, the structure of DenseNet is simpler and more effective, with fewer parameters. In ResNet, each layer is only connected to the previous layer, while in DenseNet, each layer is directly connected to all previous layers, and each layer can obtain gradients from the loss function. This can optimize the information flow and gradients of the entire network, making it easier to train and performing better on small datasets. The structure of DenseNet can achieve better feature reuse and reduce the number of parameters. Moreover, the layers of DenseNet are narrower than those of other deep learning networks (Liu et al., 2021c), making it reduce redundancy by learning with fewer feature maps. This architecture is suitable for the extraction of multi-scale landslide features under complex terrains, even with limited landslide training samples.

With the rapid expansion of deep learning methods based on CNNs, semantic segmentation models have increasingly become the standard in landslide detection. Numerous advanced semantic segmentation networks have been proposed and validated for automatic landslide detection, significantly enhancing the efficiency and accuracy of large-scale detection.

U-Net's encoder-decoder structure with skip connections has become a benchmark for landslide segmentation (Ronneberger et al., 2015). For example, Dong et al. (2022) proposed a new model, L-UNet, based on the U-Net architecture and successfully applied it to landslide extraction from remote sensing imagery. Nava et al. (2022) applied the Attention U-Net to Sentinel-1 SAR data for rapid mapping of earthquake-induced landslides, demonstrating the effectiveness of U-Net variants in pixel-level segmentation of landslide bodies under cloud-covered or topographically complex conditions.

When dealing with complex features in landslide-prone areas, DeepLab is a more suitable choice (Chen et al., 2017; Sandric et al., 2024). Built upon deep convolutional neural networks, DeepLab employs dilated convolutions to expand the receptive field and integrates an atrous spatial pyramid pooling (ASPP) module to capture multi-scale contextual information.

In contrast, the U-Net architecture is relatively simpler and better suited for small targets and high-resolution imagery, such as landslide crack segmentation or fine annotation of high-resolution UAV images. DeepLab, on the other hand, is more effective for large-scale landslide area detection and multispectral remote sensing image classification (see Fig. 2).

After achieving semantic segmentation to obtain the accurate extent of a landslide and the classification of ground objects, change detection is employed to monitor the changes in the landslide area over time. By comparing the segmentation results of multiple temporal phases or directly analyzing the feature differences, the dynamic evolution of potential hazards can be quantifie (Amankwah et al., 2022).

Wang (2023) demonstrates that 3D CNNs can directly process these 3D tensors. These models capture spatial and temporal features using convolutional kernels while transforming multi-temporal image sequences into change hotspot maps or temporal variation curves as output.

Some studies even have integrated attention mechanisms into conventional CNN architectures to enhance the analysis of multi-temporal remote sensing imagery, thereby enabling the identification of landslide hazard evolution over time.

3.2 Models for Time Series Analysis in Potential Landslide Identification

The occurrence of a landslide is a gradual accumulation process, usually influenced by a variety of factors. We refer to data that reflect the changing states of a landslide body over time as time series data. Time series data analysis aims to excavate the information hidden in the time series data to help identify potential landslides.

Different from conventional statistical or physical models, deep learning models can automatically reveal dynamic change trends and periodic patterns in the data, providing more accurate information for landslide prediction and early warning. Recently, deep learning–based temporal models, such as recurrent neural networks (RNNs), long short-term memory networks (LSTMs), and gated recurrent units (GRUs), have become key tools for extracting nonlinear dependencies and temporal evolution patterns in landslide-related time series.

Recurrent neural networks (RNNs) are a class of deep learning models specialized in processing sequential data, capable of capturing temporal dependencies within input sequences (Elman, 1990). Unlike conventional feedforward neural networks, in an RNN, each neuron not only receives the current input but also the output of the previous time step as additional input. This structure endows the RNN with a memory mechanism (Ngo et al., 2021; Zaremba et al., 2014).

In landslide prediction, RNNs have been employed to model displacement time series under rainfall or groundwater fluctuations, revealing short-term deformation patterns preceding slope failure (Chen et al., 2015; Zhang et al., 2022).

To overcome the vanishing gradient problem inherent in RNNs, LSTMs introduce memory cells and gating mechanisms that selectively retain relevant temporal information (Landi et al., 2021; Sherstinsky, 2020; Smagulova and James, 2019; Staudemeyer and Morris, 2019; Yu et al., 2019). This capability allows them to model the cumulative and delayed responses of slopes to prolonged rainfall or reservoir water level fluctuations.

LSTM models have been widely applied in landslide displacement prediction and early warning. Yang et al. (2019) analyzed the relationships among landslide deformation, rainfall, and reservoir water levels, and found that compared with static models, the LSTM approach more accurately captured the dynamic characteristics of landslides and effectively leveraged historical information. Xu and Niu (2018) used a LSTM model to predict the displacement evolution of the Baijiabao landslide using rainfall and hydrological level data, achieving a higher correlation compared with traditional regression models. In another study focused on shallow landslides, Xiao et al (2022) used a week-ahead LSTM model, which exhibited stable performance and improved prediction accuracy in short-term prediction scenarios. Additionally, Gidon et al (2023) constructed a Bi-LSTM model and achieved a detection accuracy of 93% in the Mawiongrim area. This effectively addresses the limitations of traditional methods and can provide a reliable technical solution for disaster early warning in this area as well as other similar landslide-prone areas.

The GRU is a simplified variant of the LSTM (Cho et al., 2014; Chung et al., 2014; Zhang et al., 2022b) that achieves similar accuracy with fewer parameters and reduced computational costs, making it well-suited for real-time landslide monitoring systems (Rawat et al., 2024).

Furthermore, GRU models effectively identify precursory displacement acceleration, allowing early detection of slope instability triggered by rainfall or seismic shaking (Chang et al., 2025; Yang et al., 2025).

Recently, Transformer architectures (Vaswani et al., 2017) have been introduced for time series modeling due to their ability to capture global dependencies across long sequences through the self-attention mechanism.

Unlike RNNs or LSTMs that process data sequentially, Transformers analyze all time steps in parallel, offering better scalability and modeling of long-term deformation trends.

In landslide applications, Transformer-based approaches have shown promise in integrating multi-source time series—such as rainfall, soil moisture, and deformation—into a unified temporal framework. Zhao et al. (2024) combined the strengths of CNN and Transformer architectures, selecting and analyzing nine landslide-conditioning factors to successfully achieve accurate landslide localization and detailed feature capture. Ge et al. (2024) proposed the LiteTransNet model based on the Transformer framework, effectively capturing and interpreting the varying importance of historical information during the prediction process.

Although Transformer-based models require larger training datasets and higher computational resources, their capacity to model complex, long-range dependencies and integrate multi-factor information offers significant potential for the next generation of intelligent landslide early warning systems.

In summary, RNNs and their advanced variants (LSTM, GRU) have demonstrated strong capabilities in modeling landslide time series, enabling early detection of slope deformation acceleration and rainfall-induced instability (Li et al., 2021a; Wang et al., 2020b). Transformer architectures further extend this capability to capture cross-variable and long-term dependencies (Wang et al., 2024a; Zerveas et al., 2021), offering a new direction for multi-sensor, data-driven landslide hazard prediction (see Fig. 3).

3.3 Models for Data Generation in Potential Landslide Identification

Data generation refers to modeling the underlying data distribution of data to generate entirely new samples independent of the original dataset (Kingma et al., 2014; Moreno-Barea et al., 2020; Shorten and Khoshgoftaar, 2019), thereby enriching the dataset. In potential landslide identification, data generation mitigates challenges related to data scarcity and class imbalance, which are particularly pronounced in geohazard mapping tasks where labeled landslide samples are limited. This process enhances the generalization capability of predictive models and enables the simulation of diverse landslide scenarios.

Deep generative models are the leading deep learning approach for synthetic data generation (Alam et al., 2018; Karras et al., 2020; Ma et al., 2024; Xu et al., 2015). They utilize deep neural networks to learn latent representations of data and optimize the learning process through specific objective functions. A key characteristic of deep generative models lies in their probabilistic nature—they not only classify or reconstruct data but also capture the underlying distribution of geospatial features, enabling the creation of new landslide samples that are statistically consistent with observed patterns. Commonly used deep generative models include generative adversarial networks (GANs), variational autoencoders (VAEs), and diffusion models.

GANs consist of a generator and a discriminator that compete in an adversarial process (Goodfellow et al., 2014). The generator synthesizes data resembling real samples, while the discriminator attempts to distinguish between generated and real data. Through iterative adversarial training, the generator learns to produce high-quality synthetic data that closely matches the distribution of real data (Gui et al., 2021; Saxena and Cao, 2021).

In the context of landslide studies, GANs have demonstrated strong capabilities in data augmentation and remote sensing image enhancement. For example, Al-Najjar and Pradhan (2021) proposed a novel approach that employs a GAN to generate synthetic inventory data. The results indicate that additional samples produced by the proposed GAN model can enhance the predictive performance of Decision Trees (DT), Random Forest (RF), Artificial Neural Network (ANN), and Bagging ensemble models. Feng et al. (2024) achieved the first implementation of using a GAN to generate synthetic high-quality landslide images, aiming to address the data scarcity issue that undermines the performance of landslide segmentation models.

Despite their advantages, GANs may suffer from mode collapse, leading to limited diversity in the generated data, especially when certain landslide types are underrepresented (Fang et al., 2020). Moreover, their unstable training process requires careful hyperparameter tuning and substantial computational resources, which may constrain their application in real-time hazard scenarios. Nevertheless, with improved architectures such as CGAN, Pix2Pix, and Wasserstein GAN (WGAN) ((Arjovsky et al., 2017; Isola et al., 2017; Kim and Lee, 2020; Loey et al., 2020; Qu et al., 2019; Mirza et al., 2014; Wang et al., 2019), GANs are becoming increasingly viable tools for high-resolution landslide mapping and synthetic data generation in remote sensing-based susceptibility analyses.

As a probabilistic variant of autoencoders (AEs), VAEs introduce latent-space regularization through variational inference (Kingma et al., 2013). The encoder compresses input data into a latent representation characterized by a mean and a standard deviation, while the decoder reconstructs the data by sampling from this distribution. This enables the model to generate new data with inherent randomness and diversity (Islam et al., 2021; Oliveira et al., 2022).

In landslide research, VAEs have been successfully applied to learn and reconstruct geomorphological patterns of slope instability. For instance, Cai et al. (2024) proposed and demonstrated the superior capability of the VAE-GRU model in generating narrow predictive intervals while maintaining high coverage probabilities, representing a substantial improvement over the state-of-the-art methods for probabilistic landslide prediction.

Compared with GANs, VAEs produce more diverse but slightly less detailed samples, due to their structured latent space constraints. This characteristic is particularly beneficial for exploring a wide range of potential landslide morphologies and for augmenting training datasets used in susceptibility prediction. However, VAEs may still struggle with highly imbalanced datasets, as their probabilistic reconstruction tends to favor majority classes. Integrating VAEs with stratified sampling or cost-sensitive learning could help overcome this limitation and further enhance landslide prediction performance.

When computational resources and training time permit, diffusion models provide a powerful alternative for generating high-quality, diverse, and stable data (Ho et al., 2020; Croitoru et al., 2023; Yang et al., 2023a; Zhu et al., 2023a). These models learn the data distribution by gradually adding noise to real samples (forward diffusion) and then reconstructing clean data through a reverse denoising process. The resulting models can sample new, realistic data points that reflect complex terrain and geophysical variability.

Although diffusion models are still in the exploratory phase for landslide applications, recent geospatial AI research indicates their high potential for terrain simulation and deformation modeling. Lo et al. (2024) proposed a Terrain-Feature-Guided Diffusion Model (TFDM) to fill gaps in DEM data. Similarly, Zhao et al. (2024) employed a Denoising Diffusion Probabilistic Model (DDPM) conditioned on incomplete DEMs, which serves as a transitional kernel during diffusion reversal to progressively reconstruct sharp and accurate DEM surfaces.

Deep generative models provide a transformative solution for overcoming the challenges of limited and imbalanced landslide datasets. By synthesizing realistic, diverse, and statistically consistent samples, these models can improve the robustness and generalization of landslide prediction frameworks. GANs are effective for generating visually realistic imagery and data augmentation; VAEs capture probabilistic geomorphic transitions; and diffusion models ensure stability and fidelity in high-resolution terrain synthesis.

As generative AI continues to evolve, integrating these models with multi-source remote sensing inputs and physics-based constraints holds great promise for next-generation landslide hazard identification systems. Such integration is expected to enhance data diversity, reduce labeling dependency, and enable more precise, interpretable, and generalizable predictions for landslide risk assessment and early warning.

3.4 Models for Anomaly detection in Potential Landslide Identification

Anomaly detection plays a critical role in potential landslide identification, as it enables the distinction between normal environmental variations and genuine precursors of slope instability (Deijns et al., 2020; Jiang et al., 2020). In landslide monitoring, the goal of anomaly detection is to identify subtle yet significant deviations—such as abnormal surface displacements, changes in surface coherence, or irregular sensor signals—that may occur prior to failure events. With the advancement of deep learning, data filtering has evolved from rule-based threshold detection to automated feature learning, allowing models to capture complex spatiotemporal dependencies and identify anomalies within high-dimensional, multi-source datasets.

AEs are widely used for unsupervised anomaly detection due to their ability to reconstruct input data and highlight deviations from learned normal patterns. An AE consists of an encoder that compresses data into a low-dimensional latent representation and a decoder that reconstructs it.

During training, the AE learns the intrinsic features of normal landslide data—such as sensor-based displacement time series or radar backscatter from stable slopes. When abnormal data are input, such as sudden displacement spikes or incoherent radar signals, the reconstruction error increases significantly, serving as an indicator of potential instability. For instance, Shakeel et al. (2022) developed an InSAR deformation anomaly detector based on an AE–LSTM architecture (ALADDIn). Experimental analyses using synthetic deformation test scenarios achieved an overall performance accuracy of 91.25%.

By defining a reconstruction error threshold, anomalies can be quantitatively detected. When the reconstruction error of new sensor data exceeds this threshold, it may signal slope movement acceleration or surface disturbance associated with potential landslides. Thus, AEs provide a data-driven method to detect early-warning signs without requiring manually labeled failure data.

As previously introduced, VAE a probabilistic extension of AEs. VAEs introduce stochastic latent variables characterized by mean and variance, allowing them to model data uncertainty (Kingma et al., 2013; Li et al., 2020; Park et al., 2018). During training, VAEs learn the latent distribution of normal samples and reconstruct inputs accordingly. When new observation data deviate significantly from the learned distribution, the reconstruction error increases accordingly, and this phenomenon can be used as an indicator of potential anomalies.

In landslide applications, VAEs have been shown to outperform traditional AEs in handling complex, multivariate datasets that integrate topographic, meteorological, and geotechnical factors. For example, Han et al. (2025) proposed an unsupervised failure mode recognition algorithm based on a deep convolutional autoencoder, which integrates surface displacement, vertical displacement, and rainfall monitoring data from slopes to accurately identify the developmental stages of slope failure, achieving a recognition accuracy of 99.30%.

Another study by Yadav et al. (2024) proposed a novel unsupervised change detection (CD) model, termed CLVAE, designed to learn the spatiotemporal correlations within Sentinel-1 SAR time series. The model achieved a mean IoU of 70% and a mean F1-score of 81%, outperforming comparative models by at least 6% in F1-score and 8% in IoU.

Compared to AEs, VAEs are particularly advantageous for capturing uncertainty and latent correlations between environmental variables, making them ideal for data cleaning in integrated landslide early-warning systems. However, they require larger datasets for stable training, and their probabilistic outputs may demand post-processing for operational thresholding.

GANs can also be adapted for anomaly detection by exploiting their discriminator network's ability to distinguish between real and generated data (Kang et al., 2024; Xia et al., 2022). In landslide monitoring, GAN-based anomaly detection models learn the distribution of stable slope features, and deviations from this distribution can indicate abnormal conditions (Radoi, 2022).

AnoGAN extends conventional GANs by directly incorporating data cleaning as one of its primary objectives (Lin et al., 2023; Thomine et al., 2023). It introduces an additional encoder during training, which maps input data to the latent space. The difference between this latent vector and the latent vector of normal samples generated by the generator serves as the basis for data cleaning.

RNNs and their variants are particularly effective for time series–based anomaly detection, learning temporal dependencies and predicting future trends (Zamanzadeh et al., 2024; Zhang et al., 2022a). In landslide monitoring, these models can process continuous displacement or rainfall time series to identify deviations from expected temporal behavior. These temporal models complement image-based approaches by providing continuous surveillance and early detection capabilities (Wu et al., 2024).

When combined with AEs or GANs, RNN-type architectures can form hybrid frameworks capable of both spatial and temporal anomaly detection, enabling multi-source consistency checking in landslide early-warning systems. Geiger et al (2020) demonstrated a growing trend of utilizing LSTM networks as both the generator and discriminator within GAN frameworks for time-series anomaly detection. Similarly, Whitakeret al (2023) illustrated the application of LSTM–GAN architectures in identifying temporal anomalies.

3.5 Models for Data Fusion in Potential Landslide Identification

In practical applications, the identification of potential landslide hazards is a complex task that influences by multiple factors (Zhang et al., 2018). These factors are often reflected through different data sources. We can roughly divide heterogeneous data into four categories: image data, time series data, structured data, and textual data. Data fusion is essential for the accurate identification of potential landslides. In order to better identify potential landslides, data fusion is essential.

Since heterogeneous data differ in feature scale, spatial resolution, and data modality, deep learning models are increasingly utilized to automatically extract nonlinear and high-order feature interactions across data sources, offering significant advantages over traditional statistical fusion techniques. In landslide applications, deep learning-based data fusion can integrate multi-modal inputs such as Sentinel-1 InSAR deformation, rainfall time series, and terrain derivatives for regional-scale susceptibility mapping or real-time early warning.

Due to the non-Euclidean and topologically complex nature of landslide-related terrain, conventional CNN-based models are limited in representing irregular spatial dependencies (Scarselli et al., 2008). Graph Neural Networks (GNNs) have emerged as powerful architectures to model such relationships by representing spatial entities (e.g., slope units, grid cells, or sensor nodes) as graph nodes and their geospatial or topological interactions as edges (Ying et al., 2018; Zeng et al., 2022).

In landslide identification, GNNs enable explicit modeling of spatial connectivity and geological adjacency, allowing the propagation of geomorphic and hydrological information across neighboring units. For example, Kuang et al. (2022) proposed an innovative landslide forecasting model based on Graph Neural Networks, in which graph convolutions are employed to aggregate spatial correlations among different monitoring sites. Ren et al. (2025) introduced a novel Graph Neural Network framework with conformal prediction (GNN-CF) for landslide deformation interval forecasting, addressing the limitations of traditional models in handling predictive uncertainty.

According to the differences in message passing and aggregation methods, GNNs have derived various variants. For example, graph convolutional network (GCN) is generated by generalizing the convolutional operation to graph-structured data (Kipf et al., 2016; Sharma et al., 2022; Wang et al., 2020a), and graph attention network (GAT) dynamically weights the importance of neighboring nodes by introducing the attention mechanism (Velickovic et al., 2017; Yuan et al., 2022; Zhou and

Li, 2021). The emergence of these new architectures makes GNN variants more targeted than conventional GNNs and suitable for modeling heterogeneous relationships. Currently, they are often used for weighted analysis of the impacts of different geographical factors on landslides.

Transformer architectures, characterized by the self-attention mechanism, provide another promising avenue for landslide-related data fusion (Huang and Chen, 2023; Zhao et al., 2021a). Unlike CNNs or RNNs, which process spatial or temporal sequences sequentially, Transformers can jointly capture long-range dependencies across spatial and temporal dimensions, enabling unified processing of rainfall, InSAR time series, and topographic data (Esser et al., 2021; Lv et al., 2023).

Recent studies have begun adapting Transformer variants for landslide identification. Li et al. (2023) proposed a Transformer-based deep neural network capable of identifying landslides from hillshade maps and optical imagery. Piran et al. (2024) enhanced short-term precipitation forecasting by applying transfer learning with a pre-trained Transformer model. Zhang et al. (2024) incorporated Transformer modules to build a graph-Transformer model that integrates global contextual information for the generation and analysis of landslide susceptibility maps (LSMs).

Newly Added References

Arjovsky, M., Chintala, S. and Bottou, L., 2017, July. Wasserstein GAN. arXiv preprint arXiv:1701.07875. doi:10.48550/arXiv.1701.07875.

Cai, W., Lan, F., Huang, X., Hao, J., Xia, W., Tang, R., Feng, P. and Li, H., 2024. Generative probabilistic prediction of precipitation induced landslide deformation with variational autoencoder and gated recurrent unit. Frontiers in Earth Science, 12, p.1394129. doi:10.3389/feart.2024.1394129.

Chang, F., Dong, S., Yin, H., Ye, X., Wu, Z., Zhang, W. and Zhu, H., 2025. 3D displacement time series prediction of a north-facing reservoir landslide powered by InSAR and machine learning. Journal of Rock Mechanics and Geotechnical Engineering. doi:10.1016/j.jrmge.2024.10.033.

Chen, H., Zeng, Z. and Tang, H., 2015. Landslide deformation prediction based on recurrent neural network. Neural Processing Letters, 41(2), pp.169-178. doi:10.1007/s11063-013-9318-5.

Chen, L.C., Papandreou, G., Kokkinos, I., Murphy, K. and Yuille, A.L., 2017. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. IEEE transactions on pattern analysis and machine intelligence, 40(4), pp.834-848. doi:10.1109/TPAMI.2017.2699184.

Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H. and Bengio, Y., 2014. Learning phrase representations using RNN encoder-decoder for statistical machine translation. arXiv preprint arXiv:1406.1078. doi:10.48550/arXiv.1406.1078.

Ding, X., Zhang, X., Han, J. and Ding, G., 2022. Scaling up your kernels to 31x31: Revisiting large kernel design in cnns. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 11963-11975). doi:10.48550/arXiv.2203.06717.

Ebrahimi, M.S. and Abadi, H.K., 2021, July. Study of residual networks for image recognition. In Intelligent Computing: Proceedings of the 2021 Computing Conference, Volume 2 (pp. 754-763). Cham: Springer International Publishing. doi:10.1007/978-3-030-80126-7_53.

Elman, J.L., 1990. Finding structure in time. Cognitive science, 14(2), pp.179-211. doi:10.1207/s15516709cog1402_1.

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

Geiger, A., Liu, D., Alnegheimish, S., Cuesta-Infante, A. and Veeramachaneni, K., 2020, December. Tadgan: Time series anomaly detection using generative adversarial networks. In 2020 ieee international conference on big data (big data) (pp. 33-43). IEEE. doi:10.1109/BigData50022.2020.9378139.

Hamaguchi, R., Fujita, A., Nemoto, K., Imaizumi, T. and Hikosaka, S., 2018, March. Effective use of dilated convolutions for segmenting small object instances in remote sensing imagery. In 2018 IEEE winter conference on applications of computer vision (WACV) (pp. 1442-1450). IEEE. doi:10.1109/WACV.2018.00162.

Han, N., Miao, W., Li, M., Mohamad Ismail, M.A., Hu, Q., Duan, L. and Tang, J., 2025. Integrating multi-source monitoring data and deep convolutional autoencoder technology for slope failure pattern recognition. Frontiers in Earth Science, 13, p.1531857. doi:10.3389/feart.2025.1531857.

Hasanah, S.A., Pravitasari, A.A., Abdullah, A.S., Yulita, I.N. and Asnawi, M.H., 2023. A deep learning review of resnet architecture for lung disease Identification in CXR Image. Applied sciences, 13(24), p.13111. doi:

He, K., Zhang, X., Ren, S. and Sun, J., 2016. Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778). doi:10.3390/app132413111.

Hochreiter, S. and Schmidhuber, J., 1997. Long short-term memory. Neural computation, 9(8), pp.1735-1780. doi:10.1162/neco.1997.9.8.1735.

Isola, P., Zhu, J.Y., Zhou, T. and Efros, A.A., 2017. Image-to-image translation with conditional adversarial networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1125-1134). doi:10.1109/CVPR.2017.632.

Kipf, T.N., 2016. Semi-supervised classification with graph convolutional networks. arXiv preprint arXiv:1609.02907. doi:10.48550/arXiv.1609.02907.

Kuang, P., Li, R., Huang, Y., Wu, J., Luo, X. and Zhou, F., 2022. Landslide displacement prediction via attentive graph neural network. Remote Sensing, 14(8), p.1919. doi:10.3390/rs14081919.

LeCun, Y., Bottou, L., Bengio, Y. and Haffner, P., 1998. Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11), pp.2278-2324. doi:10.1109/5.726791.

Li, J., Li, Q., Lu, J., Zheng, K., Wei, L. and Xiang, Q., 2025. A Transfer Learning Remote Sensing Landslide Image Segmentation Method Based on Nonlinear Modeling and Large Kernel Attention. Applied Sciences, 15(7), p.3855. doi:10.3390/app15073855.

Lo, K.S.H. and Peters, J., 2024. Diff-dem: A diffusion probabilistic approach to digital elevation model void filling. IEEE Geoscience and Remote Sensing Letters, 21, pp.1-5. doi:10.1109/LGRS.2024.3403835.

Mirza, M. and Osindero, S., 2014. Conditional generative adversarial nets. arXiv preprint arXiv:1411.1784. doi:10.48550/arXiv.1411.1784.

Piran, M.J., Wang, X., Kim, H.J. and Kwon, H.H., 2024. Precipitation nowcasting using transformer-based generative models and transfer learning for improved disaster preparedness. International Journal of Applied Earth Observation and Geoinformation, 132, p.103962. doi:10.1016/j.jag.2024.103962.

Rawat, P.S. and Barthwal, A., 2024. LANDSLIDE MONITOR: a real-time landslide monitoring system. Environmental Earth Sciences, 83(8), p.226. doi:10.1007/s12665-024-11526-0.

Ren, X., Liu, W., Yang, W., Mao, L. and Li, H., 2025. Landslide Deformation Uncertainty Quantification Using Conformalized Graph Neural Networks: A Case Study in Sichuan Province, China. IEEE Access. doi:10.1109/ACCESS.2025.3568273.

Scarselli, F., Gori, M., Tsoi, A.C., Hagenbuchner, M. and Monfardini, G., 2008. The graph neural network

model. IEEE transactions on neural networks, 20(1), pp.61-80. doi: 10.1109/TNN.2008.2005605.
Shakeel, A., Walters, R.J., Ebmeier, S.K. and Al Moubayed, N., 2022. ALADDIn: Autoencoder-LSTM-based anomaly detector of deformation in InSAR. IEEE Transactions on Geoscience and Remote Sensing, 60, pp.1-12. doi:10.1109/TGRS.2022.3169455.
Xiao, Y., Ju, N., He, C., Xiao, Z., & Ma, Z. (2022). Week-ahead shallow landslide displacement prediction using chaotic models and robust LSTM. Frontiers in Earth Science, 10, 965071. doi:10.3389/feart.2022.965071.
Wang, K., Wei, B., Zhao, T., Wu, G., Zhang, J., Zhu, L. and Wang, L., 2024. An Automated Approach for Mapping Mining-Induced Fissures Using CNNs and UAS Photogrammetry. Remote Sensing, 16(12), p.2090. doi:10.3390/rs16122090.
Wang, Y., Fang, Z., Wang, M., Peng, L., Hong, H., 2020b. Comparative study of landslide susceptibility mapping with different recurrent neural networks. Computers and Geosciences 138, 1-18. doi:10.1016/j.cageo.2020.104445.
Whitaker, T., 2023. LSTM-GAN for Enhanced Anomaly Detection in Time Series Data. Journal of Computer Technology and Software, 2(2).
Wu, L., Zhou, J.T., Zhang, H., Wang, S.R., Ma, T., Yan, H. and Li, S.H., 2024. Time series analysis and gated recurrent neural network model for predicting landslide displacements. Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards, 18(1), pp.172-185. doi:10.1080/17499518.2022.2138918.
Thi Ngo, P., Panahi, M., Khosravi, K., Ghorbanzadeh, O., Kariminejad, N., Cerda, A., Lee, S., 2021. Evaluation of deep learning algorithms for national scale landslide susceptibility mapping of iran. Geoscience Frontiers 12, 505-519. doi:10.1016/j.gsf.2020.06.013.
Veličković, P., Cucurull, G., Casanova, A., Romero, A., Lio, P. and Bengio, Y., 2017. Graph attention networks. arXiv preprint arXiv:1710.10903. doi:10.48550/arXiv.1710.10903.
Yadav, R., Nascetti, A., Azizpour, H. and Ban, Y., 2024. Unsupervised flood detection on SAR time series using variational autoencoder. International Journal of Applied Earth Observation and Geoinformation, 126, p.103635. doi:10.1016/j.jag.2023.103635.
Yang, H., Liu, Y., Han, Q., Xu, L., Zhang, T., Wang, Z., Yan, A., Zhao, S., Han, J. and Wang, Y., 2025. Improved Landslide Deformation Prediction Using Convolutional Neural Network–Gated Recurrent Unit and Spatial–Temporal Data. Remote Sensing, 17(4), p.727. doi:10.3390/rs17040727.
Zamanzadeh Darban, Z., Webb, G.I., Pan, S., Aggarwal, C. and Salehi, M., 2024. Deep learning for time series anomaly detection: A survey. ACM Computing Surveys, 57(1), pp.1-42. doi:10.1145/3691338.
Zhang, D., Yang, J., Li, F., Han, S., Qin, L. and Li, Q., 2022. Landslide risk prediction model using an attention-based temporal convolutional network connected to a recurrent neural network. IEEE Access, 10, pp.37635-37645. doi:10.1109/ACCESS.2022.3165051.
Zhang, Q., He, Y., Zhang, Y., Lu, J., Zhang, L., Huo, T., Tang, J., Fang, Y. and Zhang, Y., 2024. A Graph-Transformer method for landslide susceptibility mapping. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. doi:10.1109/JSTARS.2024.3437751.
Zhao, J., Yuan, Y., Dong, Y., Li, Y., Shao, C. and Yang, H., 2024. Void filling of digital elevation models based on terrain feature-guided diffusion model. Remote Sensing of Environment, 315, p.114432. doi:10.1016/j.rse.2024.114432.
Zhao, Z., Chen, T., Dou, J., Liu, G. and Plaza, A., 2024. Landslide susceptibility mapping considering landslide local-global features based on CNN and transformer. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 17, pp.7475-7489. doi:10.1109/JSTARS.2024.3379350.

Comment #5:

4. The most notable weakness of this work is the grammar and writing style, which is well below the acceptable standards of a journal paper. I found many grammar errors and several sentences which essentially repeated sentences just before them, among other issues (see below). The language was often vague, passive, and lacking focus. The authors need to carefully proofread their paper. This reads like a rough draft, not a publication-ready submission.

Response:

- We sincerely apologize for the inadequate writing quality in the original submission and fully acknowledge that the grammar and expression did not meet the high standards required for journal publication. We appreciate your detailed remarks, which helped us identify the issues more clearly. We will also take this valuable feedback as an opportunity to strengthen our own academic writing competency in future research.
- To thoroughly address this concern, we have undertaken a comprehensive revision of the entire manuscript:
- (1) We have meticulously addressed all the specific technical corrections (spelling, grammar, and wording/writing style) that you pointed out in the comments below.
- (2) Furthermore, the entire manuscript has been carefully checked and polished using advanced AI-powered writing assistance technology (specifically, Grammarly and ChatGPT) to correct grammatical errors, eliminate redundancy, and improve sentence clarity.
- (3) Following this, all authors have performed multiple rounds of manual proofreading to ensure the final text is coherent, focused, and meets the high standards expected for publication.
- We believe that the revised manuscript has been significantly improved in terms of language fluency and clarity. Once again, we are truly grateful for your valuable time and for pointing out these issues, which have been instrumental in improving our paper.

Original Description in L9

……

Next, several commonly used deep learning models are classified based on their roles in potential landslide identification, such as image analysis and processing, time series analysis.

……

Revised Description in L9

……

Next, several commonly used deep learning models are classified based on their roles in potential landslide identification, covering areas such as image analysis and time series analysis.

……

Original Description in L44-46

……

Through the training of large-scale and multi-source data, deep learning models are able to automatically extract features, capture complex nonlinear relationships, and conduct pattern recognition in high-dimensional data, which shows great potential in the identification of potential landslides (Nava et al., 2021; Yang et al., 2024c).

……

Revised Description in L44-46

……

Through training on large-scale and multi-source data, deep learning models are able to automatically extract features, capture complex nonlinear relationships, and conduct pattern recognition in high-dimensional data, which shows great potential in the identification of landslide hazards (Nava et al., 2021; Yang et al., 2024c).

……

Original Description in L86

……

Interferometric synthetic aperture radar (InSAR) has been developed based on. It obtains surface elevation information by performing coherent processing on two sets of SAR images observed in the same area (Dai et al., 2022; Ma et al., 2023b; Zeng et al., 2024).

……

Revised Description in L84

……

Interferometric synthetic aperture radar (InSAR) has been developed based on the principles of radar interferometry. It obtains surface elevation information by performing coherent processing on two sets of SAR images observed in the same area (Dai et al., 2022; Ma et al., 2023b; Zeng et al., 2024).

……

Original Description in L158

……

When equipped with LiDAR sensors, UAVs can effectively remove vegetation from the data. Then, assisting researchers to reveal landslide boundaries, crack patterns, and other deformation features hidden beneath vegetation cover.

……

Revised Description in L158

……

When equipped with LiDAR sensors, UAVs can effectively remove vegetation from the data, thereby assisting researchers to reveal landslide boundaries, crack patterns, and other deformation features hidden beneath vegetation cover.

……

Original Description in L193

……

TLS scanner can also help identify the landslide mass, that is, the flow path of the landslide materials.

……

Revised Description in L193

……

TLS can also help identify the landslide mass-that is, the flow path of the landslide materials.

……

Original Description in L218-219

……

By combining the obtained subsurface data with geomechanical equations, the position of the sliding surface or geotechnical strength parameters can be inverted.

……

Revised Description in L218-219

……

By combining the obtained subsurface data with geomechanical equations, the position of the sliding surface or geotechnical strength parameters can be inferred.

……

Original Description in L223-227

……

With continuous exploration, deep learning, through its powerful feature learning capabilities, enables the automatic extraction of meaningful features from raw data, significantly reducing manual intervention. Especially when dealing with high-dimensional and complex landslide data, deep learning models can extract deep features related to landslides from raw data in a data-driven manner, without the need for manual feature design.

……

Revised Description in L223-227

……

By leveraging its powerful feature learning capabilities, deep learning models can automatically extract salient features from complex, high-dimensional landslide data without manual design, thereby minimizing human intervention.

……

Original Description in L257-258

……

Conventional CNNs typically consist of multiple stacked convolutional layers, pooling layers, and fully connected layers. However, increasing network depth introduces challenges such as vanishing gradients and degradation arise, resulting in model performance deterioration.

……

Revised Description in L257-258

……

As the network depth increases, conventional CNN architectures may encounter issues such as vanishing gradients and performance degradation, which hinder effective feature extraction in complex landslide imagery. To address these limitations, advanced CNN variants such as ResNet, DenseNet, and U-Net have been developed to enhance feature propagation and maintain training stability.

……

Original Description in L294

……

By comparing the segmentation results of multiple temporal phases or directly analyzing the feature differences, the dynamic evolution of potential hazards can be quantifie (Amankwah et al., 2022).

……

Revised Description in L294

……

By comparing the segmentation results of multiple temporal phases or directly analyzing the feature differences, the dynamic evolution of potential hazards can be quantified (Amankwah et al., 2022).

……

Original Description in L297

……

Wang (2023) demonstrates that 3D CNNs can directly process these 3D tensors. These models capture spatial and temporal features using convolutional kernels while transforming multi-temporal image sequences into change hotspot maps or temporal variation curves as output.

……

Revised Description in L297

……

Wang (2023) demonstrates that 3D CNNs can directly process these 3D tensors. These models capture spatial and temporal features using convolutional kernels while transforming multi-temporal image sequences into change hotspot maps or temporal variation curves as outputs.

……

Original Description in L304

……

Different from conventional time series data analysis methods, using deep learning models an automatically reveal the dynamic change trends and periodic patterns in the data, providing more accurate information for landslide prediction.

……

Revised Description in L304

……

Different from conventional time series data analysis methods, using deep learning models can automatically reveal the dynamic change trends and periodic patterns in the data, providing more accurate information for landslide prediction.

……

Original Description in L315-317

……

During the training process, the RNN will process the data at each time step in sequence, continuously updating the hidden state. By combining the input 315 of the current time step with the hidden state of the previous moment for calculation to gain an understanding of the data at the current moment, this structure enables the RNN to capture the temporal evolution patterns of landslide-related factors.

……

Revised Description in L315-317

……

During the training process, the RNN processes the data at each time step in sequence, continuously updating the hidden state. By combining the input of the current time step with the hidden state of the previous moment, this structure enables the RNN to capture the temporal evolution patterns of landslide-related factors.

……

Original Description in L318-319

……

Due to conventional RNNs struggle to model long-term dependencies and limit their applicability to short-term temporal sequences, long short-term memory networks (LSTM) were developed (Wang et al., 2023b).

……

Revised Description in L318-319

……

Conventional RNNs are limited in their ability to capture long-term dependencies, which restricts their applicability to short-term temporal sequences. To overcome these limitations, long short-term memory (LSTM) networks were developed (Wang et al., 2023b).

……

Original Description in L344-345

……

Gated recurrent unit (GRU) is a simplified version of LSTM (Chung et al., 2014; Zhang et al., 2022b), which has fewer parameters. Due to their higher computational efficiency, GRU has potential advantages in real-time data processing scenarios in landslide monitoring.

GRU mainly consists of the update gate and reset gate. The update gate is used to control how much of the previous information should be preserved at the current time step, while the reset gate is used to determine whether to ignore the hidden state of the previous time step, enabling the model to adaptively learn information across different temporal scales. This dual-gate mechanism enables adaptive learning of multi-scale temporal patterns.

……

Revised Description in L344-345

……

Gated recurrent unit (GRU) is a simplified version of the LSTM with fewer parameters and higher computational efficiency, making it suitable for real-time landslide monitoring applications (Chung et al., 2014; Zhang et al., 2022b).

GRU mainly consists of an update gate and a reset gate. The update gate controls how much of the previous information is preserved at the current time step, while the reset gate determines whether to ignore the hidden state of the previous step, enabling the model to adaptively learn information across different temporal scales through this dual-gate mechanism.

……

Original Description in L348

……

GRU particularly well-suited for applications involving real-time analysis of on-site monitoring data, where rapid detection of imminent landslide risks is essential and data volume is relatively limited.

……

Revised Description in L348

……

GRU is particularly well-suited for applications involving real-time analysis of on-site monitoring data, where rapid detection of imminent landslide risks is essential and data volume is relatively limited.

……

Original Description in L352

……

Transformer was originally designed to handle sequential data in natural language processing, which was first introduced by Vaswani in 2017 (Vaswani et al., 2017). Unlike conventional recurrent and convolutional structures, the Transformer employs employs a self-attention mechanism to directly model the entire sequence.

……

Revised Description in L352

……

Transformer was originally designed to handle sequential data in natural language processing, which was first introduced by Vaswani in 2017 (Vaswani et al., 2017). Unlike conventional recurrent and convolutional structures, the Transformer employs a self-attention mechanism to directly model the entire sequence.……

Original Description in L365

……

Data generation refers to modeling the underlying data distribution of data to generate entirely new samples independent of the original dataset (Kingma et al., 2014; Moreno-Barea et al., 2020; Shorten and Khoshgoftaar, 2019), thereby enriching the dataset.

……

Revised Description in L365

……

Data generation refers to modeling the underlying data distribution to generate entirely new samples independent of the original dataset (Kingma et al., 2014; Moreno-Barea et al., 2020; Shorten and Khoshgoftaar, 2019), thereby enriching the dataset.

……

Original Description in L434-435

……

Although diffusion models demonstrate strong capabilities in generating high-quality images and handling noise, they generate superior-quality data and ensure greater training stability compared to GANs and VAEs.

……

Revised Description in L434-435

……

In addition to their strong capabilities in generating high-quality images and handling noise, diffusion models also generate superior-quality data and ensure greater training stability compared to GANs and VAEs.

……

Original Description in L448

……

The AE is a typical unsupervised learning model consisting of an encoder and a decoder The encoder compresses the input data into low-dimensional features, and then the decoder reconstructs the input.

……

Revised Description in L448

……

The AE is a typical unsupervised learning model consisting of an encoder and a decoder The encoder compresses the input data into low-dimensional features, and then the decoder reconstructs the input.

……

Original Description in L470

……

In landslide monitoring, data may be influenced by various factors, GANs demonstrate robustness by filtering out such interference, thereby enhancing data cleaning accuracy (Radoi, 2022).

……

Revised Description in L470

……

In landslide monitoring, data may be influenced by various factors. To address this, GANs demonstrate robustness by filtering out such interference, thereby enhancing data cleaning accuracy (Radoi, 2022).

……

Original Description in L485-486

……

In practical applications, the identification of potential landslide hazards is a complex task that influences by multiple factors (Zhang et al., 2018). These factors are often reflected through different data sources. We can roughly divide heterogeneous data into four categories: image data, time series data, structured data, and textual data. Data fusion is essential for the accurate identification of potential landslides. In order to better identify potential landslides, data fusion is essential.

……

Revised Description in L485-486

……

In practical applications, the identification of potential landslide hazards is a complex task that influences by multiple factors (Zhang et al., 2018). These factors are often reflected through different data sources. We can roughly divide heterogeneous data into four categories: image data, time series data, structured data, and textual data. Data fusion is essential for the accurate identification of potential landslides.

……

Original Description in L521-522

……

Transformer can also convert multimodal dFor different types of data, it transforms them into vector representations via different embedding layers.ata into a unified vector representation through different embedding layers. Subsequently, through the use of the self-attention mechanism and multilayer neural networks, these vectors are fused and feature representations are extracted, enabling the model to process and integrate data from various modalities within the same model framework (Lv et al., 2023; Tang et al., 2022).

……

Revised Description in L521-522

……

The Transformer can convert multimodal data into a unified vector representation through different embedding layers. For different types of data, it transforms them into vector representations via their respective embedding layers. Subsequently, through the use of the self-attention mechanism and multilayer neural networks, these vectors are fused and feature representations are extracted, enabling the model to process and integrate data from various modalities within the same model framework (Lv et al., 2023; Tang et al., 2022).

…….

Original Description in L542, 692, 708

……

In this section, we focus on the analysis of the second type of potential landslides. Based on triggering factors, landslides can be classified into four categories: rainfall-induced landslides, earthquake-induced landslides, human activity-induced landslides, and multi factor-induced landslides.

……

4.4 Application of Deep Learning in the Identification of Multi factor-induced Landslides

Multi factor-induced landslides result from the synergistic interaction of multiple natural and anthropogenic factors (Hao et al., 2023).

……

In multi factor-induced landslides, earthquakes and rainfall often interact with other factors.

……

With the accumulation of new data and the dynamic variations in multi factor-induced landslides, regular model updates are critical to ensuring identification accuracy and adaptability.

……

In contrast, multi factor-induced landslides necessitate models that integrate multiple triggering mechanisms and perform a comprehensive assessment of the cumulative effects of diverse contributing factors.

……

Revised Description in L542, 692, 708

……

In this section, we focus on the analysis of the second type of potential landslides. Based on triggering factors, landslides can be classified into four categories: rainfall-induced landslides, earthquake-induced landslides, human activity-induced landslides, and multi-factor-induced landslides.

……

4.4 Application of Deep Learning in the Identification of Multi-factor-induced Landslides

Multi-factor-induced landslides result from the synergistic interaction of multiple natural and anthropogenic factors (Hao et al., 2023).

……

In multi-factor-induced landslides, earthquakes and rainfall often interact with other factors.

……

With the accumulation of new data and the dynamic variations in multi-factor-induced landslides, regular model updates are critical to ensuring identification accuracy and adaptability.

……

In contrast, multi-factor-induced landslides necessitate models that integrate multiple triggering mechanisms and perform a comprehensive assessment of the cumulative effects of diverse contributing factors.

……

Original Description in L752

……

Thus, data scarcity is a common problem in the identification of potential landslide, especially in remote areas or regions with limited data accessibility.

……

Revised Description in L752

……

Thus, data scarcity is a common problem in the identification of potential landslides, especially in remote areas or regions with limited data accessibility.

…….

Original Description in L780

……

Direct fusion of such multi-modal data induces feature space incompatibility, hindering cross-modal correlation extraction (Cai et al., 2021; Jin et al., 2022).Zhang et al. (2023) highlights that even remote sensing data exhibits high heterogeneity in imaging mechanisms, illumination conditions, and spectral characteristics.

……

Revised Description in L780

……

Direct fusion of such multi-modal data induces feature space incompatibility, hindering cross-modal correlation extraction (Cai et al., 2021; Jin et al., 2022). Zhang et al. (2023) highlights that even remote sensing data exhibits high heterogeneity in imaging mechanisms, illumination conditions, and spectral characteristics.

…….

Original Description in L913

……

By fusing the static stratum data from geological surveys with the time-series data of surface deformation monitored by InSAR, the combination of static and dynamic data is realized, which can distinguish between stable slopes and areas with potential creeping deformation.

……

Revised Description in L913

……

By fusing the static stratum data from geological surveys with the time-series surface deformation data monitored by InSAR, it is possible to distinguish between stable slopes and areas with potential creeping deformation.

…….

Original Description in L928

……

Each deep learning model excels in specific tasks or data types but may underperforming in others.

……

Revised Description in L928

……

Each deep learning model excels in specific tasks or data types but may underperform in others.

…….

Original Description in L944-945

……

For example, Guo et al. (2024) utilized a stacked approach integrating a 1D-CNN, RNN, and LSTM network can form a CRNN-LSTM ensemble model.

……

Revised Description in L944-945

……

For example, Guo et al. (2024) utilized a stacked approach integrating a 1D-CNN, RNN, and LSTM network to form a CRNN-LSTM ensemble model.

…….

Original Description in L958

……

The core concept involves leverage knowledge analysis to gain a deeper understanding of landslide triggering mechanisms and mechanical behaviors, while combine data-driven methods to extract potential landslide features and patterns from monitoring data and historical records.

……

Revised Description in L958

……

The core concept involves leveraging knowledge analysis to gain a deeper understanding of landslide triggering mechanisms and mechanical behaviors, while combining data-driven methods to extract potential landslide features and patterns from monitoring data and historical records.

…….

Original Description in L1033

……

In this review, we summarized the latest advancement in the applications of deep learning models for potential landslide identification, as well as the challenges and opportunities for the future.

……

Revised Description in L1033

……

In this review, we summarized the latest advancements in the applications of deep learning models for potential landslide identification, as well as the challenges and opportunities for the future.

……

Comment #6:

5. L28: I'm not sure what you mean by the "relativity…of potential landslides". Could you clarify what is "relative" about potential landslides?

Response:

- We sincerely thank you for this insightful comment. We apologize for the lack of clarity in our original phrasing. The term "relativity" was intended to convey that the assessment of landslide potential is not absolute but is comparative and context-dependent. It refers to the relative likelihood, spatial probability, or comparative susceptibility of a landslide occurring in one area versus another, based on a set of conditioning factors (e.g., slope, geology, land use).
- We have revised the manuscript to improve precision. The phrase has been replaced with "the inherent uncertainty and dynamic nature" to better convey that landslide prediction is not absolute but is a probabilistic assessment fraught with challenges.

Original Description in Section 1

……

Due to the relativity and dynamic nature of potential landslides, the identification work becomes extremely complicated. On the one hand, it is not possible to determine that a landslide will definitely occur just because there are signs of deformation on the slope.

……

Revised Description in Section 1

……

Due to the inherent uncertainty and dynamic nature of potential landslides, the identification work becomes extremely complicated. On the one hand, it is not possible to determine that a landslide will definitely occur just because there are signs of deformation on the slope.

……

Comment #7:

6. Figure 1 contains reference to several important satellites that are not explained or mentioned anywhere in the text (Sentinel, for example). The authors should briefly describe the satellites and what kind of images they produce, either in the caption or the main body of the text, where relevant.

Response:

- Thank you for carefully pointing out this oversight! We fully agree with this comment. To address it, we have added concise descriptions of the major satellites in the caption of Figure 1. These descriptions clarify the types of data each satellite provides and their relevance to potential landslide identification.

Original Figure 1

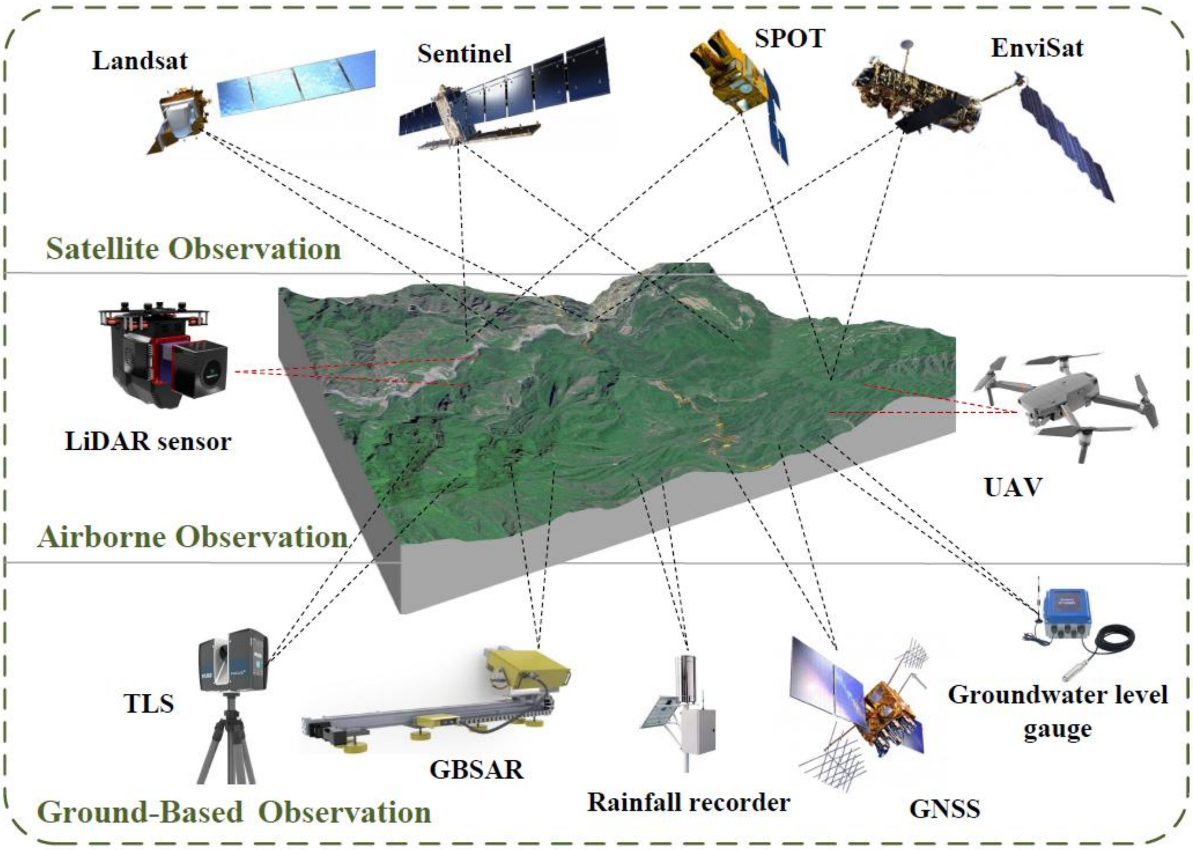

Figure 1. Data sources for potential landslide identification.

Revised Figure 1

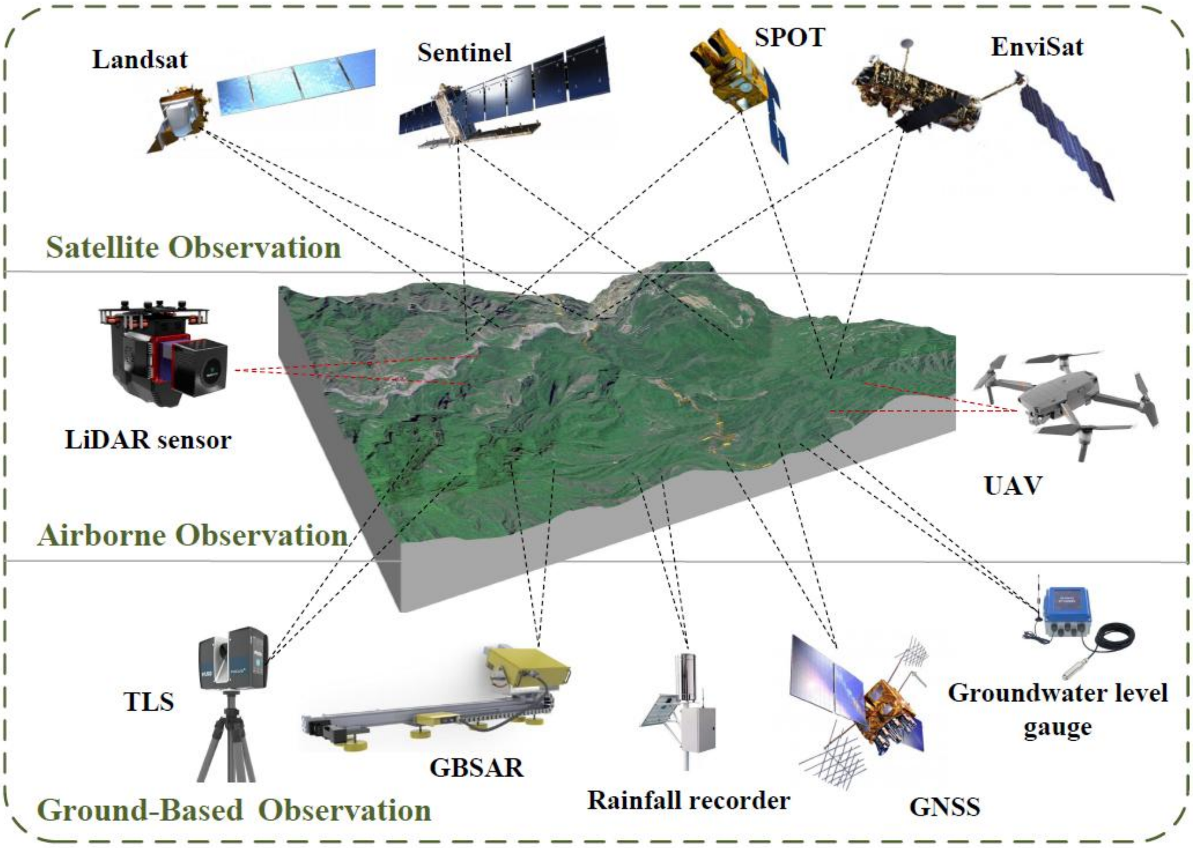

Figure 1. Data sources for potential landslide identification. Satellite observations (e.g., Landsat, Sentinel, SPOT, and Envisat) provide optical and radar imagery with varying spatial resolutions for detecting and mapping landslides. Airborne observations (LiDAR, UAV) deliver high-resolution topographic and photographic data, while ground-based observations (TLS, GBSAR, GNSS, rainfall and groundwater sensors) offer continuous in-situ monitoring of slope dynamics.

Comment #8:

7. While I appreciate the brevity of Section 1, I feel it would be improved by adding a paragraph summarizing the authors' overall takeaways and findings from this review.

Response:

- We sincerely thank you for this valuable suggestion. We have added a summary paragraph at the end of **Section 1** to clearly outline the overall takeaways and findings of this review. The added paragraph synthesizes the core contributions of our work, highlights the value of deep learning in potential landslide identification, and points out current challenges and future directions, thereby providing readers with a more complete overview.

Original Description in Section 1

……

In this review, we aim to summarize the applications of deep learning in the field of potential landslide identification, including data, models, applications, challenges, and future directions.

(1) We classify commonly used heterogeneous data into three categories for research. These data sources offer comprehensive data support for the application of deep learning in potential landslide identification.

(2) We introduce the roles of commonly used deep learning models in potential landslide identification, and compare the advantages and disadvantages among different models.

(3) We analyze the performance of deep learning models in different scenarios through case studies, discussing the adaptability of deep learning in potential landslide identification.

(4) We summarize the main challenges currently faced by the application of deep learning in potential landslide identification, and highlight new opportunities and promising future directions.

The remainder of this paper is organized as follows. Section 2 introduces seven main data sources. Section 3 summarizes five roles of deep learning models in potential landslide identification. Section 4 investigates the application of deep learning. models in four typical landslides and provides a comprehensive summary. Section 5 analyzes the current challenges in potential landslide identification. Section 6 discusses future research directions. Section 7 provides the concluding remarks.

……

Revised Description in Section 1

……

In this review, we aim to summarize the applications of deep learning in the field of potential landslide identification, including data, models, applications, challenges, and future directions.

(1) We classify commonly used heterogeneous data into three categories for research. These data sources offer comprehensive data support for the application of deep learning in potential landslide identification.

(2) We introduce the roles of commonly used deep learning models in potential landslide identification, and compare the advantages and disadvantages among different models.

(3) We analyze the performance of deep learning models in different scenarios through case studies, discussing the adaptability of deep learning in potential landslide identification.

(4) We summarize the main challenges currently faced by the application of deep learning in potential landslide identification, and highlight new opportunities and promising future directions.

Through our analysis, we identified several key trends in the application of deep learning to potential landslide identification. First, researchers are increasingly adopting multi-source data fusion techniques, integrating information from diverse sources to construct a more comprehensive representation of the geological environment (Guo et al., 2025; Liu et al., 2020; Wang et al., 2024). Second, deep learning models have been successfully applied across multiple scales, ranging from large-scale landslide susceptibility mapping with convolutional neural networks (CNNs) to real-time slope deformation monitoring with recurrent neural networks (RNNs) (Azarafza et al., 2021; Soni et al., 2025; Xie et al., 2024; Zhao et al., 2024). Despite these advances, the field continues to face critical challenges that will shape its future trajectory. The main challenges include the scarcity of high-quality, well-annotated landslide datasets; limited generalization and transferability of models across diverse geological settings; and the inherent “black box” nature of deep learning, which undermines interpretability and trust in high-stakes decision-making contexts. Consequently, future research is expected to place greater emphasis on integrating physical knowledge with data-driven approaches, thereby advancing the field from traditional, reactive post-disaster responses toward intelligent, proactive pre-disaster risk management.

……

Newly Added References

Azarafza M, Azarafza M, Akgün H, Atkinson PM, Derakhshani R. Deep learning-based landslide susceptibility mapping. Scientific reports. 2021 Dec 16;11(1):24112. doi:10.1038/s41598-021-03585-1.

Guo D, Yang X, Peng P, Zhu L, He H. The intelligent fault identification method based on multi-source information fusion and deep learning. Scientific Reports. 2025 Feb 24;15(1):6643. doi:10.1038/s41598-025-90823-5.

Liu Y, Xu C, Huang B, Ren X, Liu C, Hu B, Chen Z. Landslide displacement prediction based on multi-source data fusion and sensitivity states. Engineering Geology. 2020 Jun 20; 271:105608. doi:10.1016/j.enggeo.2020.105608.

Soni R, Alam MS, Vishwakarma GK. Prediction of InSAR deformation time-series using improved LSTM deep learning model. Scientific Reports. 2025 Feb 13;15(1):5333. doi:10.1038/s41598-024-83084-1.

Wang X, Wang D, Liu C, Zhang M, Xu L, Sun T, Li W, Cheng S, Dong J. Refined Intelligent Landslide Identification Based on Multi-Source Information Fusion. Remote Sensing. 2024 Sep 1;16(17). doi:10.3390/rs16173119.

Xie Y, Meng X, Wang J, Li H, Lu X, Ding J, Jia Y, Yang Y. Enhancing GNSS deformation monitoring forecasting with a combined VMD-CNN-LSTM deep learning model. Remote Sensing. 2024 May 16;16(10):1767. doi:10.3390/rs16101767.

Zhao Z, Chen T, Dou J, Liu G, Plaza A. Landslide susceptibility mapping considering landslide local-global features based on CNN and transformer. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. 2024 Mar 19; 17:7475-89. doi:10.1109/JSTARS.2024.3379350.

Comment #9:

8. Section 2.1.1: I would recommend the section on SAR also discuss NISAR. This is quite timely; the satellite just launched, is the most expensive earth-imaging satellite ever, and part of its mission is to monitor and better understand natural processes on Earth. Further, the authors should mention that another benefit of SAR is that it can image Earth regardless of illumination (ie day or night) and weather conditions (eg cloudy), which is not true of optical remote sensing.

Response:

- Thank you for your excellent and constructive suggestions. We fully agree that discussing the recently launched NISAR mission is highly timely and relevant for our review, as it represents a significant advancement in SAR for Earth observation. We also appreciate the comment regarding the all-weather capability of SAR, which is indeed a critical advantage over optical imaging and we apologize for this omission in our original manuscript.

- According to your reasonable suggestion, we have revised the relevant content. (Please see revised **Subsection 2.2.1** for details).
-

Original Description in Subsection 2.1.1

……

SAR is an active microwave remote sensing system. It is not only capable of acquiring data on demand by actively emitting microwave signals but also facilitates partial penetration of vegetation cover through its longer wavelength bands (such as the L-band), thereby allowing the retrieval of surface deformation information beneath vegetated areas. The time series data provided by SAR can serve as input for deep learning models, allowing these models to be trained to identify long-term patterns of terrain change. Continuous monitoring of potential landslide areas is crucial, and SAR is widely employed in high-risk environments.

……

Revised Description in Subsection 2.1.1

……

SAR is an active microwave remote sensing system. It is not only capable of acquiring data on demand by actively emitting microwave signals but also facilitates partial penetration of vegetation cover through its longer wavelength bands (such as the L-band), thereby allowing the retrieval of surface deformation information beneath vegetated areas.

A critical operational advantage of SAR lies in its capacity to image regardless of illumination (day or night) and weather conditions (e.g., cloud cover) (Koukiou, 2024). The continuous, unimpeded time series data this provides is essential for serving as input to deep learning models, allowing these models to be trained to identify long-term patterns of terrain change. For this reason, SAR is widely employed for the crucial task of continuous monitoring in high-risk environments, where cloud cover and the timing of a disaster are unpredictable.

Notably, the NASA–ISRO SAR Mission (NISAR), jointly developed by the National Aeronautics and Space Administration (NASA) and the Indian Space Research Organisation (ISRO), was successfully launched in 2025 (ISRO, 2025; NASA Science, 2025). The satellite carries both L-band and S-band SAR systems, enabling more precise and frequent measurements of surface deformation. With a revisit period of approximately 12 days, it delivers globally consistent coverage with a balanced spatial and temporal resolution. This capability provides researchers with abundant and continuous observations, supporting large-scale, high spatiotemporal resolution landslide early detection and dynamic monitoring.

……

Newly Added References

ISRO (2025). NISAR – NASA–ISRO Synthetic Aperture Radar Mission. Indian Space Research Organisation. Retrieved October 26, 2025, from https://www.isro.gov.in/Mission_GSLVF16_NISAR_Home.html.

Koukiou, G., 2024. SAR Features and Techniques for Urban Planning—A Review. Remote Sensing, 16(11), p.1923. doi:10.3390/rs16111923.

NASA Science (2025). Mission overview – NASA–ISRO Synthetic Aperture Radar (NISAR). National Aeronautics and Space Administration. Retrieved October 26, 2025, from https://science.nasa.gov/mission/nisar/mission-overview/.

Comment #10:

9. Section 2.1.2 would benefit from more citations to previous work. In particular, it needs more citations to support its statements. This also benefits readers who want to learn more. For example "Its application in geological hazard investigations dates back to the 1970s" and "currently capable of achieving spatial resolutions as fine as 0.3 meters or better" are both claims that have no supporting citation.

Response:

- Thank you for your valuable comment! We agree entirely that providing appropriate citations strengthens the arguments in this section and provides valuable resources for readers. We have now carefully revised **Subsection 2.1.2** and added several key references to support the claims, particularly the two specific examples mentioned by you. (Please see new **Subsection 2.1.2** for details).

Original Description in Subsection 2.1.2

……

Optical remote sensing refers to the acquisition of surface information through sensors that measure reflected solar radiation. Its application in geological hazard investigations dates back to the 1970s.

Optical remote sensing offers high resolution, currently capable of achieving spatial resolutions as fine as 0.3 meters or better. In potential landslide identification, it not only facilitates the retrieval of detailed surface textures and color characteristics using rich spectral data but also enables the direct identification of morphological features and object contours via visual interpretation of imagery (Cheng and Han, 2016; Li et al., 2022b).

Landslide formation typically follows a progressive process from deformation to failure, accompanied by precursor indicators such as tensile cracks, stepped scarps, and localized collapses. These indicators exhibit distinct spectral signatures in optical imagery compared to their surroundings, enabling both manual interpretation and automated detection.

In vegetated mountainous regions, surface vegetation often undergoes detectable changes before a landslide event. Optical remote sensing leverages multispectral data, particularly red and near-infrared bands, to monitor vegetation health and identify potential landslide zones. Furthermore, the calculation of the normalized difference vegetation index (NDVI) facilitates the evaluation of vegetation health in potential landslide regions, providing critical insights into potential landslide precursors (Verrelst et al., 2015)

……

Revised Description in Subsection 2.1.2

……

Optical remote sensing refers to the acquisition of surface information through sensors that measure reflected solar radiation. Its application in geological hazard investigations dates back to the 1970s (Fu et al., 2024; Liu and Wu, 2016).

Optical remote sensing offers high resolution, currently capable of achieving spatial resolutions as fine as 0.3 meters or better. For example, Maxar's WorldView-3 delivers 0.31 m panchromatic imagery (Hu et al., 2016; Longbotham et al., 2014), while India's Cartosat-3 satellite achieves panchromatic imagery with a resolution of up to 0.25 m (Gupta et al., 2024). In potential landslide identification, it not only facilitates the retrieval of detailed surface textures and color characteristics using rich spectral data but also enables the direct identification of morphological features and object contours via visual interpretation of imagery (Cheng and Han, 2016; Li et al., 2022b; Ma and Wang, 2023).

Landslide formation typically follows a progressive process from deformation to failure, accompanied by precursor indicators such as tensile cracks, stepped scarps, and localized collapses. These indicators exhibit distinct spectral signatures in optical imagery compared to their surroundings, enabling both manual interpretation and automated detection.

In vegetated mountainous regions, surface vegetation often undergoes detectable changes before a landslide event. Optical remote sensing leverages multispectral data, particularly red and near-infrared bands, to monitor vegetation health and identify potential landslide zones (Coluzzi et al., 2025). Furthermore, the calculation of the normalized difference vegetation index (NDVI) facilitates the evaluation of vegetation health in potential landslide regions, providing critical insights into potential landslide precursors (Verrelst et al., 2015).

……

Newly Added References

Coluzzi R, Perrone A, Samela C, Imbrenda V, Manfreda S, Pace L, Lanfredi M. Rapid landslide detection from free optical satellite imagery using a robust change detection technique. Scientific Reports. 2025 Feb 8;15(1):4697. doi:10.1038/s41598-025-89542-8.

Fu S, de Jong SM, Hou X, de Vries J, Deijns A, de Haas T. A landslide dating framework using a combination of Sentinel-1 SAR and-2 optical imagery. Engineering Geology. 2024 Feb 1; 329:107388. doi: 10.1016/j.enggeo.2023.107388.

Gupta A, Paul S, Bhattacharya A, Jain P. A framework for realistic paired dataset generation for deep learning based restoration of satellite images. InIGARSS 2024-2024 IEEE International Geoscience and Remote Sensing Symposium 2024 Jul 7 (pp. 6997-7002). IEEE. doi:10.1109/IGARSS53475.2024.10640440.

Hu F, Gao XM, Li GY, Li M. DEM extraction from worldview-3 stereo-images and accuracy evaluation. The

International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. 2016 Jun 3;41:327-32. doi:10.5194/isprs-archives-XLI-B1-327-2016.
Liu Y, Wu L. Geological disaster recognition on optical remote sensing images using deep learning. Procedia Computer Science. 2016 Jan 1; 91:566-75. doi:10.1016/j.procs.2016.07.144.
Longbotham N, Pacifici F, Baugh B, Camps-Valls G. Prelaunch assessment of worldview-3 information content. In2014 6th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS) 2014 Jun 24 (pp. 1-4). IEEE. doi:10.1109/WHISPERS.2014.8077566.
Ma H, Wang F. Factors controlling the formation and movement of clustered shallow landslides triggered by the extreme rainstorm in July 2023 in Beijing, China. Geomorphology. 2025 Jun 1; 478:109728. doi: 10.1016/j.geomorph.2025.109728.

Comment #11:

10. Section 2.1.2: I would recommend more detailed discussion of multi- and hyper-spectral images and their application to landslides. You briefly mention it but I feel that more discussion is warranted given their prevalence in earth monitoring (especially via deep learning).

Response:

- We are grateful to you for this excellent suggestion. We completely agree that a more in-depth discussion on multi- and hyperspectral imagery is crucial, especially given their growing importance in landslide monitoring and the role of deep learning in automating their analysis. This expansion undoubtedly strengthens the manuscript and provides readers with a more comprehensive overview.
- In response to this comment, we have significantly expanded **Subsection 2.1.2** by adding a dedicated paragraph that further explains the characteristics and advantages of multi- and hyperspectral data and their applications in traditional machine learning and deep learning approaches (Please see new **Subsection 2.1.2** for details).

Original Description in Subsection 2.1.2

……

In vegetated mountainous regions, surface vegetation often undergoes detectable changes before a landslide event. Optical remote sensing leverages multispectral data, particularly red and near-infrared bands, to monitor vegetation health and identify potential landslide zones. Furthermore, the calculation of the normalized difference vegetation index (NDVI) facilitates the evaluation of vegetation health in potential landslide regions, providing critical insights into potential landslide precursors (Verrelst et al., 2015).

……

Revised Description in Subsection 2.1.2

……

In vegetated mountainous regions, surface vegetation often undergoes detectable changes before a landslide event. Optical remote sensing leverages multispectral data, particularly red and near-infrared bands, to monitor vegetation health and identify potential landslide zones. Furthermore, the calculation of the normalized difference vegetation index (NDVI) facilitates the evaluation of vegetation health in potential landslide regions, providing critical insights into potential landslide precursors (Verrelst et al., 2015).

However, the broad spectral bands of multispectral sensors limit their ability to detect more subtle, diagnostically specific precursory signals. The advancement beyond broad-band multispectral imaging to hyperspectral imaging has opened new avenues for landslide precursor detection (Kilgore and Restrepo, 2025; Ye et al., 2019). Hyperspectral sensors capture hundreds of contiguous spectral bands, enabling the identification of specific mineralogies (e.g., expansive clays like smectite that influence slope stability) and subtle geochemical alterations on slope surfaces. For instance, the shifting absorption features in the short-wave infrared (SWIR) region can signal changes in soil water content and mineral composition that often precede failure (Thimsen et al., 2017).

The integration of these rich spectral datasets with deep learning architectures has significantly advanced automated landslide analysis (Huang et al., 2022; Shahabi et al., 2021). These models excel at learning complex patterns from high-dimensional spectral-spatial information, enabling highly accurate detection of landslide scars and even precursory features like cracks and seepage zones that are otherwise challenging to identify.

……

Newly Added References

Huang L, Luo R, Liu X, Hao X. Spectral imaging with deep learning. Light: Science & Applications. 2022 Mar 16;11(1):61. doi:10.1038/s41377-022-00743-6.

Kilgore A, Restrepo C. Integrating Hyperspectral Imaging, Plant Functional Diversity, and Soil-Lithology to Uncover Mountainscape Disturbance Dynamics Induced by Landsliding. Remote Sensing. 2025 May 22;17(11):1806. doi:10.3390/rs17111806.

Shahabi H, Rahimzad M, Tavakkoli Piralilou S, Ghorbanzadeh O, Homayouni S, Blaschke T, Lim S, Ghamisi P. Unsupervised deep learning for landslide detection from multispectral sentinel-2 imagery. Remote Sensing. 2021 Nov 20;13(22):4698. doi:10.3390/rs13224698.

Thimsen E, Sadtler B, Berezin MY. Shortwave-infrared (SWIR) emitters for biological imaging: a review of challenges and opportunities. Nanophotonics. 2017 Aug 28;6(5):1043-54. doi:10.1515/nanoph-2017-0039.

Ye C, Li Y, Cui P, Liang L, Pirasteh S, Marcato J, Goncalves WN, Li J. Landslide detection of hyperspectral remote sensing data based on deep learning with constrains. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. 2019 Nov 25;12(12):5047-60. doi:10.1109/JSTARS.2019.2951725.

Comment #12:

11. Section 2 would benefit from a more discussion comparing and contrasting these different data sources. Lines 173-176 do a good job of this with SAR and GB-SAR. More discussion similar to this for other methods would improve this section, in my opinion.

Response:

- We sincerely appreciate your constructive and helpful comments!
- Your suggestion regarding enhancing the discussion by comparing and contrasting the different data sources in Section 2 is highly valuable and crucial for improving the depth and clarity of this section. We fully agree with your assessment and have thoroughly revised the manuscript based on your guidance.
- Following your suggestions, we have enhanced **Section 2** by adding comparative discussions at three main levels:
- (1) Added comparison within **Subsection 2.1**: A new paragraph has been added following the descriptions of space-borne SAR and optical remote sensing. This paragraph explicitly compares and contrasts these two technologies in terms of their operating principles, key strengths, limitations, and their complementary roles in regional-scale landslide identification.
- (2) Added comparison within **Subsection 2.2**: A new comparative summary paragraph has been added after the introductions of airborne LiDAR and UAV. This section now contrasts the applicability, cost-effectiveness, and operational complexity of airborne LiDAR and UAVs, framing them as a bridge between satellite and ground-based data.
- (3) Enhanced comparison within **Subsection 2.3**: The discussion comparing GB-SAR, TLS, and ground-based sensor devices has been significantly deepened.
- Please see the revised **Section 2** for details.
-
- We sincerely appreciate the time and effort you dedicated to reviewing our manuscript and providing these constructive comments, which have significantly improved the quality of our paper.

Original Description in Section 2

2.1 Satellite Observation Data

2.1.1 Space-borne SAR

……

2.1.2 Optical Remote Sensing

……

Landslide formation typically follows a progressive process from deformation to failure, accompanied by precursor indicators such as tensile cracks, stepped scarps, and localized collapses. These indicators exhibit distinct spectral signatures in optical imagery compared to their surroundings, enabling both manual interpretation and automated detection.

In vegetated mountainous regions, surface vegetation often undergoes detectable changes before a landslide event. Optical remote sensing leverages multispectral data, particularly red and near-infrared bands, to monitor vegetation health and identify potential landslide zones. Furthermore, the calculation of the normalized difference vegetation index (NDVI) facilitates the evaluation of vegetation health in potential landslide regions, providing critical insights into potential landslide precursors (Verrelst et al., 2015).

2.2 Airborne Remote Sensing Data

2.2.1 Airborne Light Detection and Ranging (LiDAR)

……

2.2.2 Unmanned Aerial Vehicle (UAV)

……

With the rapid advancement of UAVs, centimeter-level vertical and oblique aerial photogrammetry is now achievable (Fan et al., 2020). The high-definition cameras mounted on UAVs are able to capture the subtle cracks on the surface of the mountain.

These cracks may be early signs of a landslide (Sun et al., 2024a). By conducting a comparative analysis of the images taken at different times, the development and changes of the cracks can be monitored, including the increase in the length, width and depth of the cracks, as well as the changes in the crack orientation. In some mountainous areas or valleys, there may be a large number of loose accumulations. These accumulations may trigger landslides under specific conditions. Aerial photography by UAVs can clearly identify information such as the distribution range, accumulation quantity and accumulation shape of these loose accumulations, and assess their potential threats to the surrounding environment.

When equipped with LiDAR sensors, UAVs can effectively remove vegetation from the data. Then, assisting researchers to reveal landslide boundaries, crack patterns, and other deformation features hidden beneath vegetation cover. This integrated approach combines the strengths of photogrammetry and LiDAR, allowing for rapid deployment and targeted area monitoring while mitigating the challenges posed by vegetation cover in landslide detection and assessment.

After extreme weather events such as heavy rainstorms or geological events like earthquakes occur, the stability of the mountain may be affected, making it prone to triggering geological hazards. UAVs even can quickly conduct aerial monitoring of the relevant areas after extreme weather.

2.3 Ground-based Observation Data

2.3.1 Ground-based Synthetic Aperture Radar (GB-SAR)

……

2.3.2 Terrestrial Laser Scanning (TLS)

……

By quickly and massively collecting spatial point position information, TLS can automatically splice and rapidly obtain the appearance of the measured object. It can be used to construct high-precision surface models and appearance models of buildings and structures. The 3D model can display the shape and structure of the mountain and the detailed features of the ground surface from different angles and in all directions (Zhou et al., 2024), enabling geological experts and engineers to have a more intuitive understanding of the overall situation of the landslide area. For example, the cracks in the mountain, the loose accumulations, and the degree of weathering of the rocks can be clearly seen, providing richer information for the identification of potential landslide hazards.

Currently, TLS is commonly used in critical areas requiring localized precision. For historical landslide masses, it captures reactivation indicators such as rear tensile cracks and frontal bulging, with data input into anomaly detection models to identify reactivation signals.

…...

Revised Description in Section 2

2.1 Satellite Observation Data

2.1.1 Space-borne SAR

……

2.1.2 Optical Remote Sensing

……

Landslide formation typically follows a progressive process from deformation to failure, accompanied by precursor indicators such as tensile cracks, stepped scarps, and localized collapses. These indicators exhibit distinct spectral signatures in optical imagery compared to their surroundings, enabling both manual interpretation and automated detection.

In vegetated mountainous regions, surface vegetation often undergoes detectable changes before a landslide event. Optical remote sensing leverages multispectral data, particularly red and near-infrared bands, to monitor vegetation health and identify potential landslide zones (Fiorucci et al., 2018). Furthermore, the calculation of the normalized difference vegetation index (NDVI) facilitates the evaluation of vegetation health in potential landslide regions, providing critical insights into potential landslide precursors (Verrelst et al., 2015).

While both space-borne SAR and optical remote sensing are pivotal for large-area landslide screening, they offer complementary capabilities and have distinct limitations. Optical remote sensing provides intuitive visual interpretation of geomorphological features but is rendered useless by cloud cover and darkness. In contrast, space-borne SAR, with its all-weather, day-and-night imaging capability, excels in detecting millimeter-to-centimeter-scale surface deformation through InSAR techniques, which is a direct precursor to landslide failure. However, InSAR performance can be degraded in heavily vegetated areas due to temporal decorrelation and in steep terrain due to geometric distortions (Lin et al., 2022; Yan et al., 2024), areas where optical stereo imaging for DEM generation might be less affected. Therefore, the integration of SAR-derived deformation maps and optical-based geomorphological maps is considered a best practice for regional-scale landslide inventory mapping and preliminary hazard assessment (Xun et al., 2022).

2.2 Airborne Remote Sensing Data

2.2.1 Airborne Light Detection and Ranging (LiDAR)

……

2.2.2 Unmanned Aerial Vehicle (UAV)

……

With the rapid advancement of UAVs, centimeter-level vertical and oblique aerial photogrammetry is now achievable (Fan et al., 2020). The high-definition cameras mounted on UAVs are able to capture the subtle cracks on the surface of the mountain.

These cracks may be early signs of a landslide (Sun et al., 2024a). By conducting a comparative analysis of the images taken at different times, the development and changes of the cracks can be monitored, including the increase in the length, width and depth of the cracks, as well as the changes in the crack orientation. In some mountainous areas or valleys, there may be a large number of loose accumulations. These accumulations may trigger landslides under specific conditions. Aerial photography by UAVs can clearly identify information such as the distribution range, accumulation quantity and accumulation shape of these loose accumulations, and assess their potential threats to the surrounding environment.

Airborne platforms bridge the gap between satellite and ground-based observations. Airborne LiDAR is unparalleled in generating high-precision DEM, revealing concealed paleo-landslides and subtle topographic features critical for hazard mapping. However, its deployment is costly and logistically complex. UAVs, as a flexible and cost-effective alternative, have democratized high-resolution data acquisition. They can be equipped with various sensors (e.g., optical, multispectral, and even lightweight LiDAR) to conduct rapid response surveys following triggering events such as earthquakes or heavy rainfall (Han et al., 2023; Sestras et al., 2025). While UAV-derived models have ultra-high resolution, their coverage is limited per sortie compared to airborne campaigns. The choice between them often involves a trade-off between coverage, cost, operational flexibility, and the specific requirement for vegetation penetration.

2.3 Ground-based Observation Data

2.3.1 Ground-based Synthetic Aperture Radar (GB-SAR)

……

2.3.2 Terrestrial Laser Scanning (TLS)

……

By quickly and massively collecting spatial point position information, TLS can automatically splice and rapidly obtain the appearance of the measured object. It can be used to construct high-precision surface models and appearance models of buildings and structures. The 3D model can display the shape and structure of the mountain and the detailed features of the ground surface from different angles and in all directions (Zhou et al., 2024), enabling geological experts and engineers to have a more intuitive understanding of the overall situation of the landslide area. For example, the cracks in the mountain, the loose accumulations, and the degree of weathering of the rocks can be clearly seen, providing richer information for the identification of potential landslide hazards.

Currently, TLS is commonly used in critical areas requiring localized precision (Abellán et al., 2009; Teng et al., 2022). For historical landslide masses, it captures reactivation indicators such as rear tensile cracks and frontal bulging, with data input into anomaly detection models to identify reactivation signals.

Ground-based techniques provide the highest precision for monitoring a specific slope of interest. GB-SAR and TLS are both non-contact remote sensing methods, but they operate on different principles. GB-SAR offers continuous, all-weather, mm-level deformation monitoring over a large area (several km^2) from a single station, making it ideal for early warning. Its drawback is the need for a stable, opposing installation point with a clear line-of-sight (Monserrat et al., 2013). TLS, on the other hand, provides mm-to-cm-level 3D point clouds of the slope surface, excellent for quantifying volume changes and detailed geometric changes. However, it is typically used for periodic surveys rather than continuous monitoring and has occlusion shadows (Huang et al., 2019).

……

Newly Added References

Abellán, A., Jaboyedoff, M., Oppikofer, T. and Vilaplana, J.M., 2009. Detection of millimetric deformation using a terrestrial laser scanner: experiment and application to a rockfall event. Natural Hazards and Earth System Sciences, 9(2), pp.365-372. doi:10.5194/nhess-9-365-2009.

Fiorucci, F., Giordan, D., Santangelo, M., Dutto, F., Rossi, M. and Guzzetti, F., 2018. Criteria for the optimal selection of remote sensing optical images to map event landslides. Natural Hazards and Earth System Sciences, 18(1), pp.405-417. doi:10.17605/OSF.IO/GD2U9.

Han, L., Duan, P., Liu, J. and Li, J., 2023. Research on landslide trace recognition by fusing UAV-based lidar dem multi-feature information. Remote Sensing, 15(19), p.4755. doi:10.3390/rs15194755.

Huang, R., Jiang, L., Shen, X., Dong, Z., Zhou, Q., Yang, B. and Wang, H., 2019. An efficient method of monitoring slow-moving landslides with long-range terrestrial laser scanning: a case study of the Dashu landslide in the Three Gorges Reservoir Region, China. Landslides, 16(4), pp.839-855. doi:10.1007/s10346-018-1118-6.

Lin, Y.N., Chen, Y.C., Kuo, Y.T. and Chao, W.A., 2022. Performance study of landslide detection using multi-temporal SAR images. Remote Sensing, 14(10), p.2444. doi:10.3390/rs14102444.

Monserrat, O., Moya, J., Luzi, G., Crosetto, M., Gili, J.A. and Corominas, J., 2013. Non-interferometric GB-SAR measurement: application to the Vallcebre landslide (eastern Pyrenees, Spain). Natural Hazards and Earth System Sciences, 13(7), pp.1873-1887. doi:10.5194/nhess-13-1873-2013.

Sestras, P., Badea, G., Badea, A.C., Salagean, T., Oniga, V.E., Roşca, S., Bilaşco, Ş., Bruma, S., Spalević, V., Kader, S. and Billi, P., 2025. A novel method for landslide deformation monitoring by fusing UAV photogrammetry and LiDAR data based on each sensor's mapping advantage in regards to terrain feature. Engineering Geology, 346, p.107890. doi:10.1016/j.enggeo.2024.107890.

Teng, J., Shi, Y., Wang, H. and Wu, J., 2022. Review on the research and applications of TLS in ground surface and constructions deformation monitoring. Sensors, 22(23), p.9179. doi:10.3390/s22239179.

Xun, Z., Zhao, C., Kang, Y., Liu, X., Liu, Y. and Du, C., 2022. Automatic extraction of potential landslides by integrating an optical remote sensing image with an InSAR-derived deformation map. Remote Sensing, 14(11), p.2669. doi:10.3390/rs14112669.

Yan, L., Xiong, Q., Li, D., Cheon, E., She, X. and Yang, S., 2024. InSAR-Driven Dynamic Landslide Hazard Mapping in Highly Vegetated Area. Remote Sensing, 16(17), p.3229. doi:10.3390/rs16173229.

Comment #13:

12. L167: "However, due to the influence of various factors, the identification results may not always be fully accurate, leading to potential misjudgments." As a review article, your job is to tell the reader what the “various factors” are! You should elaborate and cite sources about what can cause inaccuracies in identification.

Response:

- We sincerely appreciate your constructive and helpful comments! We agree that a more detailed explanation of the factors affecting identification accuracy is necessary in a review paper. We have now elaborated on the specific factors that can lead to inaccuracies in landslide identification using remote sensing and have supported these points with relevant literature citations. The revised text now reads. Please see the revised **Subsection 2.3** for all details. Thanks.

Original Description in Subsection 2.3

……

Satellite observation and airborne remote sensing are mainly employed for identifying potential landslides based on surface morphology. However, due to the influence of various factors, the identification results may not always be fully accurate, leading to potential misjudgments. Therefore, the potential landslide points identified through remote sensing still necessitate field investigations by researchers for verification, differentiation, confirmation, or exclusion of hazards. In some cases, additional. on-site observation and monitoring methods are needed for accurate assessment. Commonly used ground-based monitoring methods include ground-based SAR, 3D laser scanners and various sensor devices deployed or installed on the ground.

……

Revised Description in Subsection 2.3

……

Satellite observation and airborne remote sensing are mainly employed for identifying potential landslides based on surface morphology. However, these approaches are often affected by vegetation cover, viewing geometry, and atmospheric noise, which may lead to misclassification or omission (Almalki et al., 2022; Dubovik et al., 2021). Therefore, the potential landslide points identified through remote sensing still necessitate field investigations by researchers for verification, differentiation, confirmation, or exclusion of hazards. In some cases, additional. on-site observation and monitoring methods are needed for accurate assessment. Commonly used ground-based monitoring methods include ground-based SAR, 3D laser scanners and various sensor devices deployed or installed on the ground.

……

Newly Added References

Almalki R, Khaki M, Saco PM, Rodriguez JF. Monitoring and mapping vegetation cover changes in arid and semi-arid areas using remote sensing technology: A review. Remote Sensing. 2022 Oct 14;14(20):5143. doi:10.3390/rs14205143.

Dubovik O, Schuster GL, Xu F, Hu Y, Bösch H, Landgraf J, Li Z. Grand challenges in satellite remote sensing. Frontiers in Remote Sensing. 2021 Feb 24; 2:619818. doi:10.3389/frsen.2021.619818.

Comment #14:

13. Figure 2(b) and 2(c) are nice illustrations. However, it is not clear to me Figure 2(a) is trying to convey, particularly the top and bottom panels.

Response:

- Thank you for pointing this out! We agree that Figure 2(a) was not sufficiently clear in its current form. Our intention was to illustrate three typical CNN-based applications in landslide recognition: (i) feature extraction, where CNNs learn to highlight potential landslide-prone areas from raw imagery; (ii) semantic segmentation, where CNNs classify each pixel as landslide or non-landslide; and (iii) change detection, where pre- and post-event images are compared to detect newly emerged or expanded landslides.
- To avoid potential misunderstanding, we have revised Figure 2(a) to consistently display the input (pre-identification image) and output (post-identification result) representations. **Please see new Figure 2**.

Original Figure 2

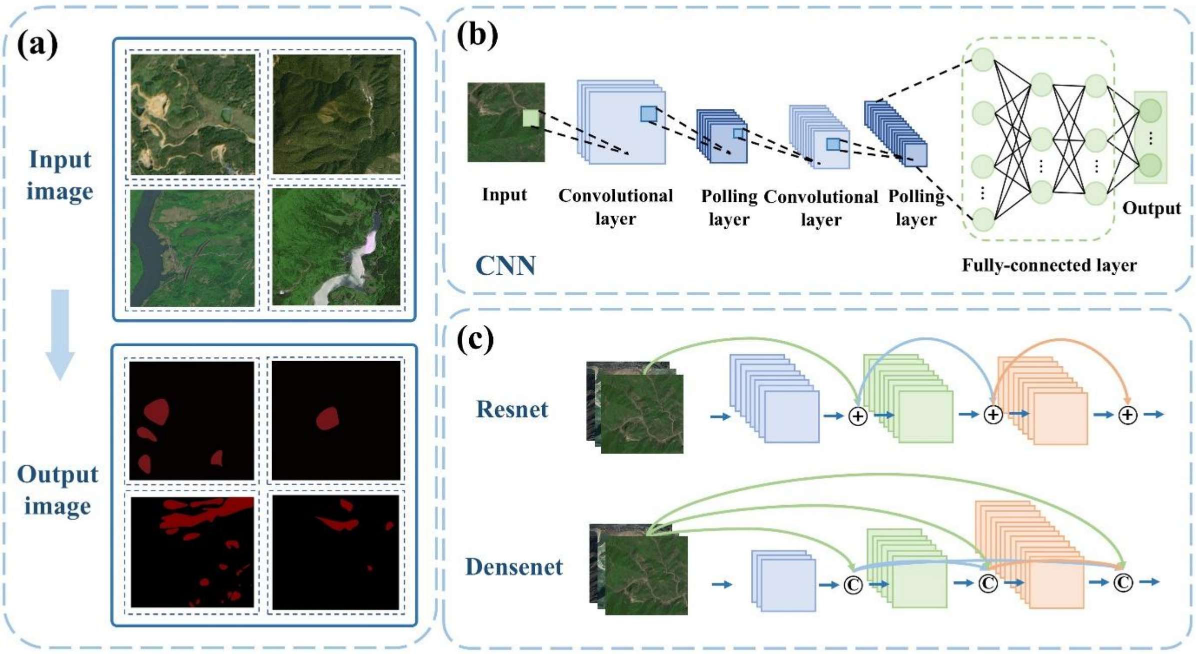

Figure 2. The role of deep learning models in image analysis and processing. (a) Three commonly used applications of CNNs in image processing for potential landslide identification. (b) Schematic of a basic CNN architecture. A conventional CNN typically comprises stacked convolutional layers, pooling layers, and fully connected layers. (c) Comparative schematic of ResNet and DenseNet architectures. In contrast to ResNet, which combines features through summation before passing them to subsequent layers, DenseNet integrates features via channelwise concatenation.

Revised Figure 2

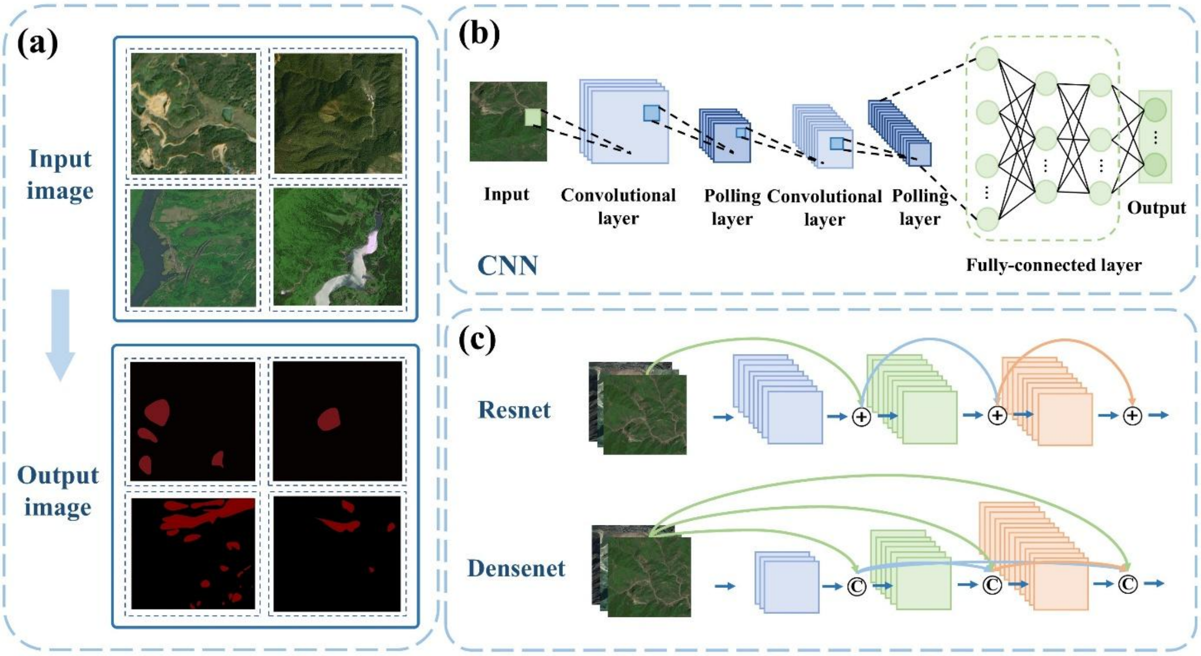

Figure 2. The role of deep learning models in image analysis and processing. (a) Comparison of landslide images before and after identification. (b) Schematic of a basic CNN architecture. A conventional CNN typically comprises stacked convolutional layers, pooling layers, and fully connected layers. (c) Comparative schematic of ResNet and DenseNet architectures. In contrast to ResNet, which combines features through summation before passing them to subsequent layers, DenseNet integrates features via channelwise concatenation.

Comment #15:

14. With regards to data, are there any benchmark datasets for landslide identification? If so, which models/methods are state-of-the-art on these datasets? this would be worthwhile to discuss, and if not, a strong recommendation to the community would be to construct such benchmark datasets to encourage further deep learning research in the area.

33. Section 5.1 makes good points, but the concept of limited training data should be discussed earlier, for example when discussing data in section 2.

Response:

- We sincerely appreciate your two highly insightful comments! You have correctly highlighted that the issue of "data" represents one of the core challenges in this field, and we acknowledge that our discussion of this aspect could indeed be presented earlier and in a more systematic manner. 14, concerning benchmark datasets, is particularly critical, while 33 has been instrumental in helping us refine the logical structure of the manuscript.

- To address both comments simultaneously, we have decided to relocate the discussion on limited training data from **Subsection 5.1** to the newly added **Subsection 2.4**. In **Subsection 2.4**, we also introduce the commonly used datasets for landslide identification. In addition, in **Section 5** and **Subsection 6.1**, we will strongly recommend that the research community collaboratively establish standardized benchmark datasets to further advance deep learning research in this field.

  The revisions are as follows:
- Several benchmark datasets for landslide detection have recently been released and are now widely used in the community. Representative examples include the **CAS Landslide Dataset**, **Landslide4Sense (L4S) benchmark**, **Diverse Mountainous Landslide Dataset (DMLD) and slope-unit-based benchmark datasets.**
- On these datasets, state-of-the-art methods typically include advanced segmentation architectures such as Transformer-based models (e.g., Swin Transformer, SegFormer) as well as improved U-Net variants, often combined with multi-source data fusion strategies and self-training techniques.
- While these datasets represent major progress, remaining gaps include (1) broader geographical diversity with held-out test regions, (2) standardized instance-level annotations (polygons vs. coarse masks), and (3) common evaluation protocols (splits and metrics).

Added Description in Subsection 2.4

2.4 Summary of Data Source for Potential Landslide Identification

In summary, no single data source is sufficient for a comprehensive landslide hazard identification framework. Regional-scale satellite data, particularly InSAR, is optimal for the early detection of pre-landslide deformations over vast areas. Airborne platforms, such as UAVs, then provide high-resolution optical and LiDAR data to characterize the precise geometry and activity of identified potential landslides. Finally, ground-based and in-situ sensors enable site-specific, real-time monitoring of high-risk slopes, validating remote sensing findings and supporting early warning systems. The strategic integration of these multi-platform data is crucial for transitioning from regional screening to mechanistic understanding and risk mitigation.

Beyond these general data modalities, recent years have also witnessed the emergence of benchmark datasets that serve as standardized testbeds for developing and evaluating deep learning methods in landslide identification. Such datasets are essential for ensuring reproducibility, enabling fair comparison across models, and accelerating methodological advances. Representative examples include the CAS Landslide Dataset, a large-scale, multi-sensor dataset explicitly designed for deep-learning-based landslide mapping (Xu et al., 2024); the Landslide4Sense (L4S) benchmark, developed within an international competition, which provides multisource satellite image patches (Ghorbanzadeh et al., 2022); and the Diverse Mountainous Landslide Dataset (DMLD), which emphasizes high-resolution instances from complex mountainous terrains (Chen et al., 2024). In addition, slope-unit-based benchmark datasets have been constructed to support susceptibility mapping and regional-scale comparisons (Martinello et al., 2021).

Overall, these datasets serve as valuable resources for pixel-level segmentation and slope-unit-based susceptibility modeling. However, in practice, the compilation of landslide inventories faces considerable challenges, making it difficult to obtain comprehensive and accurate records (Kong et al., 2025; Lee et al., 2018). Consequently, data scarcity remains a common issue in landslide hazard identification, particularly in remote regions or areas with limited accessibility. Therefore, it is necessary to further expand their geographical coverage and establish standardized evaluation protocols.

Newly Added References

Chen, J., Zeng, X., Zhu, J., Guo, Y., Hong, L., Deng, M. and Chen, K., 2024. The diverse mountainous landslide dataset (DMLD): a high-resolution remote sensing landslide dataset in diverse mountainous regions. Remote Sensing, 16(11), p.1886. doi:10.3390/rs16111886.

Ghorbanzadeh, O., Xu, Y., Zhao, H., Wang, J., Zhong, Y., Zhao, D., Zang, Q., Wang, S., Zhang, F., Shi, Y. and Zhu, X.X., 2022. The outcome of the 2022 landslide4sense competition: Advanced landslide detection from multisource satellite imagery. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 15, pp.9927-9942. doi:10.1109/JSTARS.2022.3220845.

Martinello, C., Cappadonia, C., Conoscenti, C., Agnesi, V. and Rotigliano, E., 2021. Optimal slope units partitioning in landslide susceptibility mapping. Journal of Maps, 17(3), pp.152-162. doi:10.1080/17445647.2020.1805807.

Xu, Y., Ouyang, C., Xu, Q., Wang, D., Zhao, B. and Luo, Y., 2024. Cas landslide dataset: A large-scale and multisensor dataset for deep learning-based landslide detection. Scientific Data, 11(1), p.12. doi:10.1038/s41597-023-02847-z.

Original Description in Section 5

5 Deep Learning for Potential Landslide Identification: Challenges

5.1 Data Quality and Availability

In potential landslide identification, the performance of deep learning models is critically dependent on both data quality and availability (Alzubaidi et al., 2023; Gaidzik and Ramirez-Herrera, 2021; Whang et al., 2023). Low-quality or unreliable data directly impair the models' feature extraction capabilities, while insufficient data availability constrains their generalization capacity and real-time monitoring efficacy (Azarafza et al., 2021; Xiao and Zhang, 2023).

In reality, the collection of landslide inventories faces many difficulties and it is hard to obtain them comprehensively and accurately. Thus, data scarcity is a common problem in the identification of potential landslide, especially in remote areas or regions with limited data accessibility. In such cases, deep learning models may suffer from overfitting or insufficient generalization ability due to a lack of samples (Kong et al., 2025; Lee et al., 2018). Although there are large-scale datasets such as the CAS landslide dataset, they are still insufficient compared with the data requirements of deep learning models (Xu et al., 2024).

In the natural environment, non-landslide states are the norm, while the landslide state is relatively rare. This leads to the data collected mainly consisting of normal geological conditions, with much less data representing potential landslides. Such a severe skewness in the class distribution results in a serious imbalance in the data, that is, there is a huge difference in quantity between the minority class (landslide samples) and the majority class (non-landslide samples) (Jiang et al., 2024). Gupta and Shukla (2023) demonstrated that this data imbalance can cause learning algorithms to be biased towards the majority class, perform poorly on the minority class. This bias impedes the predictive ability of the learning algorithms, and ultimately lead to the final model's poor performance in identifying and predicting the minority class of landslide samples.

……

5.2 Limitations of Deep Learning Models

……

Since the probability values output by the model lack physical significance, they cannot reflect geological uncertainties. In practical applications, it often happens that the model's prediction of high-risk areas may not distinguish between the "uncertainty caused by data absence" and the "risk of the geological conditions themselves". Even geological experts struggle to verify the rationality of these features, thereby hindering the adoption of model results in practical engineering applications.

In addition, there is also a certain contradiction between the data-driven feature learning exhibited by deep learning models and the complexity of the real world. This is because the models tend to capture the statistical patterns on the surface of the data rather than the physical mechanisms that are universal across different fields. However, the natural environment is characterized by infinite diversity, dynamism, and uncertainty. In the identification of potential landslides, this may lead to the need for repeatedly investing a large amount of annotation costs when deploying across regions and different sensors.

……

Revised Description in Section 5

5 Deep Learning for Potential Landslide Identification: Challenges

5.1 Data Quality and Availability

In potential landslide identification, the performance of deep learning models is critically dependent on both data quality and availability (Alzubaidi et al., 2023; Gaidzik and Ramirez-Herrera, 2021; Whang et al., 2023). Low-quality or unreliable data directly impair the models' feature extraction capabilities, while insufficient data availability constrains their generalization capacity and real-time monitoring efficacy (Azarafza et al., 2021; Xiao and Zhang, 2023).

In the natural environment, non-landslide states are the norm, while the landslide state is relatively rare. This leads to the data collected mainly consisting of normal geological conditions, with much less data representing potential landslides. Such a severe skewness in the class distribution results in a serious imbalance in the data, that is, there is a huge difference in quantity between the minority class (landslide samples) and the majority class (non-landslide samples) (Jiang et al., 2024). Gupta and Shukla (2023) demonstrated that this data imbalance can cause learning algorithms to be biased towards the majority class, perform poorly on the minority class. This bias impedes the predictive ability of the learning algorithms, and ultimately lead to the final model's poor performance in identifying and predicting the minority class of landslide samples.

……

5.2 Limitations of Deep Learning Models

……

Since the probability values output by the model lack physical significance, they cannot reflect geological uncertainties. In practical applications, it often happens that the model's prediction of high-risk areas may not distinguish between the "uncertainty caused by data absence" and the "risk of the geological conditions themselves" (Achu et al., 2023; Feng et al., 2022). Even geological experts struggle to verify the rationality of these features, thereby hindering the adoption of model results in practical engineering applications.

There is also a certain contradiction between the data-driven feature learning exhibited by deep learning models and the complexity of the real world. This is because the models tend to capture the statistical patterns on the surface of the data rather than the physical mechanisms that are universal across different fields. However, the natural environment is characterized by infinite diversity, dynamism, and uncertainty. In the identification of potential landslides, this may lead to the need for repeatedly investing a large amount of annotation costs when deploying across regions and different sensors.

In addition, although various types of data are available, the absence of standardized datasets with high-quality annotations has severely hindered the development and fair comparison of deep learning models (Fang et al., 2024). Existing models are often trained and evaluated on independent, task-specific datasets, which prevents an objective assessment of state-of-the-art performance and limits our ability to evaluate their true generalization capacity across different regions and triggering factors.

……

Newly Added References

Achu, A.L., Aju, C.D., Di Napoli, M., Prakash, P., Gopinath, G., Shaji, E. and Chandra, V., 2023. Machine-learning based landslide susceptibility modelling with emphasis on uncertainty analysis. Geoscience Frontiers, 14(6), p.101657. doi:10.1016/j.gsf.2023.101657.

Fang, C., Fan, X., Wang, X., Nava, L., Zhong, H., Dong, X., Qi, J. and Catani, F., 2024. A globally distributed dataset of coseismic landslide mapping via multi-source high-resolution remote sensing images. Earth System Science Data, 16(10), pp.4817-4842. doi:10.5194/essd-16-4817-2024.

Feng, H., Miao, Z. and Hu, Q., 2022. Study on the uncertainty of machine learning model for earthquake-induced landslide susceptibility assessment. Remote Sensing, 14(13), p.2968. doi:10.3390/rs14132968.

Original Description in Subsection 6.1

……

The combination of multi-source data fusion and deep learning is essentially a deep coupling of data advantages and model advantages (Chen et al., 2023a; Zheng et al., 2021). The former fills information gaps and reduces uncertainties by integrating diverse heterogeneous data, while the latter unleashes the potential of data through automated feature engineering and nonlinear modeling. This integration not only improves the accuracy of potential landslide identification but also drives the paradigm shift of geological hazard monitoring from experience-driven to data intelligence-driven. In the future, with the development of cross-modal pre-trained models and edge intelligence technologies, the collaboration between multi-source data fusion and deep learning will demonstrate greater application value in fields such as real-time early warning and hazard simulation, becoming a core technological engine for building an integrated "aerial-space-ground-subsurface" monitoring system.

……

Revised Description in Subsection 6.1

……

The combination of multi-source data fusion and deep learning is essentially a deep coupling of data advantages and model advantages (Chen et al., 2023a; Zheng et al., 2021). The former fills information gaps and reduces uncertainties by integrating diverse heterogeneous data, while the latter unleashes the potential of data through automated feature engineering and nonlinear modeling. This integration not only improves the accuracy of potential landslide identification but also drives the paradigm shift of geological hazard monitoring from experience-driven to data intelligence-driven. In the future, with the development of cross-modal pre-trained models and edge intelligence technologies, the collaboration between multi-source data fusion and deep learning will demonstrate greater application value in fields such as real-time early warning and hazard simulation, becoming a core technological engine for building an integrated "aerial-space-ground-subsurface" monitoring system.

To truly leverage the integrated monitoring method, we strongly advocate for the community-driven development of a benchmark that intrinsically incorporates this multi-modal philosophy. An ideal benchmark would not only include optical imagery but also co-registered data from SAR, LiDAR, DEM, and even ground-based sensor time series, reflecting the integrated monitoring reality. Establishing such a benchmark is the essential next step to translate our data fusion capabilities into reliable, reproducible, and advanced AI solutions for global landslide risk reduction.

……

Comment #16:

15. Throughout Section 3, when introducing a new architecture, the authors should specifically cite the paper that introduced that architecture. This both credits the original work and provides the reader with references to important papers. For example, ResNet should be attributed to Kaiming He et al., GCN to Kipf and Welling, and so on.

28. L505: Should cite the original GCN paper by Kipf and Welling

Response:

- We are grateful for your constructive feedback in identifying this shortcoming, which has helped us improve the clarity and rigor of the manuscript. We fully agree with your observation and apologize for the omission of these references in the original manuscript. In line with your suggestions, we have thoroughly revised **Section 3** (as well as other relevant parts of the manuscript) and incorporated the original seminal references for all newly introduced model architectures.

Original Description in Section 3

3 Deep Learning for Potential Landslide Identification: Models

3.1 Models for Image Analysis and Processing in Potential Landslide Identification

……

Convolutional neural networks (CNNs) represent the fundamental architecture in image processing. A CNN primarily comprises convolutional layers, pooling layers, and fully connected layers, each performing predefined functions on its input data (Kattenborn et al., 2021; Liu et al., 2022a).

……

ResNet addresses these limitations by integrating residual blocks into the foundational CNN framework (Qi et al., 2020; Yang et al., 2022). These residual blocks utilize shortcut connections that preserve original feature information. This framework facilitates the construction of ultra-deep networks capable of extracting high-level semantic features for landslide detection, thereby enhancing adaptability to complex terrain classification tasks (Ullo et al., 2021).

……

DenseNet is a further innovation of ResNet (Huang et al., 2017). Both of these neural networks are based on a similar idea, which is to establish a "shortcut" between different layers. However, the structure of DenseNet is simpler and more effective, with fewer parameters.

……

When dealing with complex features in landslide-prone areas, DeepLab is a more suitable choice (Sandric et al., 2024). Built upon deep convolutional neural networks, DeepLab employs dilated convolutions to expand the receptive field and integrates an atrous spatial pyramid pooling (ASPP) module to capture multi-scale contextual information.

……

3.2 Models for Time Series Analysis in Potential Landslide Identification

……

Recurrent neural networks (RNNs) are a class of deep learning models specialized in processing sequential data, capable of capturing temporal dependencies within input sequences (Ngo et al., 2021; Zaremba et al., 2014). Unlike conventional feedforward neural networks, in an RNN, each neuron not only receives the current input but also the output of the previous time step as additional input. This structure endows the RNN with a memory mechanism.

……

LSTM is an enhancement of RNNs, primarily processing long sequence data. Compared to standard RNNs, the hidden layer architecture of LSTM is much more complex. By incorporating memory cells and gating mechanisms, LSTM selectively propagates critical information across multiple time steps, thereby effectively capturing long-range temporal dependencies (Landi et al., 2021; Yu et al., 2019).

……

Gated recurrent unit (GRU) is a simplified version of LSTM (Chung et al., 2014; Zhang et al., 2022b), which has fewer parameters. Due to their higher computational efficiency, GRU has potential advantages in real-time data processing scenarios in landslide monitoring.

……

3.3 Models for Data Generation in Potential Landslide Identification

……

With the proposal and development of GANs, researchers have introduced various enhanced structures that are more effectively applied to potential landslide identification. For example, the conditional GAN (cGAN) (Kim and Lee, 2020; Loey et al., 202), Pix2Pix (Qu et al., 2019), and Wasserstein GAN (WGAN) (Wang et al., 2019).

……

3.4 Models for Data Cleaning in Potential Landslide Identification

……

3.5 Models for Data Fusion in Potential Landslide Identification

……

Due to the complex non-Euclidean structural characteristics of the geological environment, topographic data and their spatial relationships related to landslide hazards, conventional methods such as CNNs have difficulty in handling these relationships. As a neural network architecture for processing graph-structured data, graph neural networks (GNNs) can effectively model such spatial relationships (Ying et al., 2018; Zeng et al., 2022). They can treat the nodes in the geographical space (such as different geographical location points) and their connection relationships (such as the distance between adjacent nodes, terrain undulations, etc.) as the structure of a graph for processing.

……

Revised Description in Section 3

3 Deep Learning for Potential Landslide Identification: Models

3.1 Models for Image Analysis and Processing in Potential Landslide Identification

……

Convolutional neural networks (CNNs) represent the fundamental architecture in image processing (LeCun et al., 1998). A CNN primarily comprises convolutional layers, pooling layers, and fully connected layers, each performing predefined functions on its input data (Kattenborn et al., 2021; Liu et al., 2022a).

……

ResNet addresses these limitations by integrating residual blocks into the foundational CNN framework (He et al., 2016; Qi et al., 2020; Yang et al., 2022). These residual blocks utilize shortcut connections that preserve original feature information. This framework facilitates the construction of ultra-deep networks capable of extracting high-level semantic features for landslide detection, thereby enhancing adaptability to complex terrain classification tasks (Ullo et al., 2021).

……

DenseNet is a further innovation of ResNet (Huang et al., 2017). Both of these neural networks are based on a similar idea, which is to establish a "shortcut" between different layers. However, the structure of DenseNet is simpler and more effective, with fewer parameters.

……

When dealing with complex features in landslide-prone areas, DeepLab is a more suitable choice (Chen et al., 2017; Sandric et al., 2024). Built upon deep convolutional neural networks, DeepLab employs dilated convolutions to expand the receptive field and integrates an atrous spatial pyramid pooling (ASPP) module to capture multi-scale contextual information.

……

3.2 Models for Time Series Analysis in Potential Landslide Identification

……

Recurrent neural networks (RNNs) are a class of deep learning models specialized in processing sequential data, capable of capturing temporal dependencies within input sequences (Elman, 1990). Unlike conventional feedforward neural networks, in an RNN, each neuron not only receives the current input but also the output of the previous time step as additional input. This structure endows the RNN with a memory mechanism (Ngo et al., 2021; Zaremba et al., 2014).

……

LSTM is an enhancement of RNNs, primarily processing long sequence data (Hochreiter and Schmidhuber, 1997). Compared to standard RNNs, the hidden layer architecture of LSTM is much more complex. By incorporating memory cells and gating mechanisms, LSTM selectively propagates critical information across multiple time steps, thereby effectively capturing long-range temporal dependencies (Landi et al., 2021; Yu et al., 2019).

……

Gated recurrent unit (GRU) is a simplified version of LSTM, which has fewer parameters (Cho et al., 2014). Due to their higher computational efficiency, GRU has potential advantages in real-time data processing scenarios in landslide monitoring (Chung et al., 2014; Zhang et al., 2022b).

……

3.3 Models for Data Generation in Potential Landslide Identification

……

With the proposal and development of GANs, researchers have introduced various enhanced structures that are more effectively applied to potential landslide identification. For example, the conditional GAN (cGAN) (Kim and Lee, 2020; Loey et al., 2020; Mirza et al., 2014), Pix2Pix (Isola et al., 2017; Qu et al., 2019), and Wasserstein GAN (WGAN) (Arjovsky et al., 2017; Wang et al., 2019).

……

3.4 Models for Data Cleaning in Potential Landslide Identification

……

3.5 Models for Data Fusion in Potential Landslide Identification

……

Due to the complex non-Euclidean structural characteristics of the geological environment, topographic data and their spatial relationships related to landslide hazards, conventional methods such as CNNs have difficulty in handling these relationships. As a neural network architecture for processing graph-structured data, graph neural networks (GNNs) can effectively model such spatial relationships (Scarselli et al., 2008). They can treat the nodes in the geographical space (such as different geographical location points) and their connection relationships (such as the distance between adjacent nodes, terrain undulations, etc.) as the structure of a graph for processing (Ying et al., 2018; Zeng et al., 2022).

……

According to the differences in message passing and aggregation methods, GNNs have derived various variants. For example, graph convolutional network (GCN) is generated by generalizing the convolutional operation to graph-structured data (Kipf et al., 2016; Sharma et al., 2022; Wang et al., 2020a), and graph attention network (GAT) dynamically weights the importance of neighboring nodes by introducing the attention mechanism (Velickovic et al., 2017; Yuan et al., 2022; Zhou and Li, 2021). The emergence of these new architectures makes GNN variants more targeted than conventional GNNs and suitable for modeling heterogeneous relationships. Currently, they are often used for weighted analysis of the impacts of different geographical factors on landslides.

……

Newly Added References

Arjovsky, M., Chintala, S. and Bottou, L., 2017, July. Wasserstein GAN. arXiv preprint arXiv:1701.07875. doi:10.48550/arXiv.1701.07875.

Chen, L.C., Papandreou, G., Kokkinos, I., Murphy, K. and Yuille, A.L., 2017. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. IEEE transactions on pattern analysis and machine intelligence, 40(4), pp.834-848. doi:10.1109/TPAMI.2017.2699184.

Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H. and Bengio, Y., 2014. Learning phrase representations using RNN encoder-decoder for statistical machine translation. arXiv preprint arXiv:1406.1078. doi:10.48550/arXiv.1406.1078.

Elman, J.L., 1990. Finding structure in time. Cognitive science, 14(2), pp.179-211. doi:10.1207/s15516709cog1402_1.

He, K., Zhang, X., Ren, S. and Sun, J., 2016. Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778). doi:10.1007/978-3-319-24574-4_28.

Hochreiter, S. and Schmidhuber, J., 1997. Long short-term memory. Neural computation, 9(8), pp.1735-1780. doi:10.1162/neco.1997.9.8.1735.

Isola, P., Zhu, J.Y., Zhou, T. and Efros, A.A., 2017. Image-to-image translation with conditional adversarial networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1125-1134). doi:10.1109/CVPR.2017.632.

Kipf, T.N., 2016. Semi-supervised classification with graph convolutional networks. arXiv preprint arXiv:1609.02907. doi:10.48550/arXiv.1609.02907.

LeCun, Y., Bottou, L., Bengio, Y. and Haffner, P., 1998. Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11), pp.2278-2324. doi:10.1109/5.726791.

Mirza, M. and Osindero, S., 2014. Conditional generative adversarial nets. arXiv preprint arXiv:1411.1784. doi:10.48550/arXiv.1411.1784.

Scarselli, F., Gori, M., Tsoi, A.C., Hagenbuchner, M. and Monfardini, G., 2008. The graph neural network model. IEEE transactions on neural networks, 20(1), pp.61-80. doi: 10.1109/TNN.2008.2005605.

Veličković, P., Cucurull, G., Casanova, A., Romero, A., Lio, P. and Bengio, Y., 2017. Graph attention networks. arXiv preprint arXiv:1710.10903. doi:10.48550/arXiv.1710.10903.

Comment #17:

16. L273: What does “feature reuse” mean? Potentially rephrase or clarify.

Response:

- Thank you for pointing out the misleading part. To clarify, use the term "feature reuse" to refer to the phenomenon where feature maps (the output of layers) from earlier in the network are directly used as input for multiple subsequent layers. This allows the network to preserve and leverage low-level or intermediate features throughout the depth of the network, improving efficiency and reducing the need to re-learn redundant features.
- We apologize for the overly general statement, which may have caused confusion for readers. Following your suggestion, we have clarified this term in the revised manuscript.

Original Description in Subsection 3.1

……

DenseNet is a further innovation of ResNet. Both of these neural networks are based on a similar idea, which is to establish a "shortcut" between different layers. However, the structure of DenseNet is simpler and more effective, with fewer parameters. In ResNet, each layer is only connected to the previous layer, while in DenseNet, each layer is directly connected to all previous layers, and each layer can obtain gradients from the loss function. This can optimize the information flow and gradients of the entire network, making it easier to train and performing better on small datasets. The structure of DenseNet can achieve better feature reuse and reduce the number of parameters. Moreover, the layers of DenseNet are narrower than those of other deep learning networks, making it reduce redundancy by learning with fewer feature maps. This architecture is suitable for the extraction of multi-scale landslide features under complex terrains, even with limited landslide training samples.

……

Revised Description in Subsection 3.1

……

DenseNet is a further innovation of ResNet. Both of these neural networks are based on a similar idea, which is to establish a "shortcut" between different layers. However, the structure of DenseNet is simpler and more effective, with fewer parameters. In ResNet, each layer is only connected to the previous layer, while in DenseNet, each layer is directly connected to all previous layers, and each layer can obtain gradients from the loss function. This can optimize the information flow and gradients of the entire network, making it easier to train and performing better on small datasets. The structure of DenseNet enables more effective reuse of features, meaning that each layer can directly access and build upon the feature maps generated by all preceding layers instead of re-learning similar representations. This dense connectivity not only strengthens information and gradient flow across the network but also reduces redundancy and the total number of parameters. Moreover, the layers of DenseNet are narrower than those of other deep learning networks, making it reduce redundancy by learning with fewer feature maps. This architecture is suitable for the extraction of multi-scale landslide features under complex terrains, even with limited landslide training samples.

……

Comment #18:

17. L276: "...even with limited landslide training samples". This claim is not supported by a citation. These large networks typically need a lot of training samples. Is there previous work that successfully applied ResNet or DenseNet with limited training samples and was successful? If so, how much training data did they use?

Response:

- We appreciate your observation regarding the insufficient justification of DenseNet/ResNet effectiveness under limited sample conditions. We fully agree that large neural networks generally require substantial training data. However, in practice, even with relatively small sample sizes (e.g., a few hundred or fewer than one hundred images), architectures such as ResNet and DenseNet can still achieve satisfactory performance when combined with appropriate strategies such as pretraining, data augmentation, and model fine-tuning. To support this point and directly address your comment, we have reviewed and incorporated relevant literature and added the corresponding references to the manuscript.
- Li et al., (2021) trained ResNet, DenseNet, and other models with 100, 1,000, and 10,000 samples. The results demonstrated that when the sample size reached approximately 1,000, the performance of DenseNet was already very close to that achieved with 10,000 samples. This is because increasing the number of layers in ResNet and DenseNet can enhance their performance, whereas fewer layers can yield better results for VGG and U-Net models. Considering the cost of sample generation, it is recommended to use approximately 1,000 samples for deep learning–based

landslide detection.

- Ullo et al., (2021) achieved high performance using only 160 images (landslide and non-landslide) by combining data augmentation and fine-tuning, with ResNet-101 as the backbone. The reported results were precision equals to 1.00, recall 0.93, and F1 measure 0.97.
- Cai et al., (2021) constructed a "landslide sample library" (landslide and non-landslide) in the Three Gorges Reservoir area. Although the study did not explicitly report experiments under very small training sets, the use of data augmentation and fine-tuning enabled DenseNet to perform well even when the dataset was not extremely large. While the sample size was not very small (i.e., not reduced to only tens or hundreds of images), the "sample library + augmentation + fine-tuning" strategy demonstrated that DenseNet remains effective with moderately sized datasets (e.g., several thousand to tens of thousands of samples).
- To bolster the credibility of this claim, we have added the specific sample-size comparisons from the two aforementioned studies to the relevant paragraphs of the manuscript. (Please see the new **Subsection 3.1** for details)
- We thank you again for your detailed suggestions, which have enabled us to strengthen the rigor of our conclusions.

Original Description in Subsection 3.1

……

DenseNet is a further innovation of ResNet (Huang et al., 2017). Both of these neural networks are based on a similar idea, which is to establish a "shortcut" between different layers. However, the structure of DenseNet is simpler and more effective, with fewer parameters. In ResNet, each layer is only connected to the previous layer, while in DenseNet, each layer is directly connected to all previous layers, and each layer can obtain gradients from the loss function. This can optimize the information flow and gradients of the entire network, making it easier to train and performing better on small datasets. The structure of DenseNet can achieve better feature reuse and reduce the number of parameters. Moreover, the layers of DenseNet are narrower than those of other deep learning networks (Liu et al., 2021c), making it reduce redundancy by learning with fewer feature maps. This architecture is suitable for the extraction of multi-scale landslide features under complex terrains, even with limited landslide training samples.

……

Revised Description in Subsection 3.1

……

DenseNet is a further innovation of ResNet (Huang et al., 2017). Both of these neural networks are based on a similar idea, which is to establish a "shortcut" between different layers. However, the structure of DenseNet is simpler and more effective, with fewer parameters. In ResNet, each layer is only connected to the previous layer, while in DenseNet, each layer is directly connected to all previous layers, and each layer can obtain gradients from the loss function. This can optimize the information flow and gradients of the entire network, making it easier to train and performing better on small datasets. The structure of DenseNet can achieve better feature reuse and reduce the number of parameters. Moreover, the layers of DenseNet are narrower than those of other deep learning networks (Liu et al., 2021c), making it reduce redundancy by learning with fewer feature maps. This architecture is suitable for the extraction of multi-scale landslide features under complex terrains, even with limited landslide training samples (Cai et al., 2021; Li et al., 2021; Ullo et al., 2021).

……

Newly Added References

Cai, H., Chen, T., Niu, R. and Plaza, A., 2021. Landslide detection using densely connected convolutional networks and environmental conditions. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 14, pp.5235-5247. doi:10.1109/JSTARS.2021.3079196.

Li, C., Yi, B., Gao, P., Li, H., Sun, J., Chen, X. and Zhong, C., 2021. Valuable clues for DCNN-based landslide detection from a comparative assessment in the Wenchuan earthquake area. Sensors, 21(15), p.5191. doi:10.3390/s21155191.

Ullo, S.L., Mohan, A., Sebastianelli, A., Ahamed, S.E., Kumar, B., Dwivedi, R. and Sinha, G.R., 2021. A new mask R-CNN-based method for improved landslide detection. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 14, pp.3799-3810. doi:10.1109/JSTARS.2021.3064981.

Comment #19:

18. L277: “semantic segmentation” should be defined - it is an important concept that an ML novice may not know the meaning of.

Response:

- Thank you for pointing out the details of our paper that lack clarity. This is really helpful. Following your suggestion, we have added a definition of semantic segmentation in **Subsection 3.1.**

Original Description in Subsection 3.1

……

With the rapid expansion of deep learning methods based on CNNs, semantic segmentation models have increasingly become the standard in landslide detection. Numerous advanced semantic segmentation networks have been proposed and validated for automatic landslide detection, significantly enhancing the efficiency and accuracy of large-scale detection. U-Net is a typical example (Ronneberger et al., 2015), which features a U-shaped architecture. U-Net employs an encoder-decoder structure, where the encoder is similar to conventional CNNs, progressively reducing image resolution and extracting features through convolution and pooling operations; the decoder then restores the image resolution through transposed convolution or upsampling operations (Dong et al., 2022; Nava et al., 2022). Skip connections bridge low-level detail features with deep semantic features, thereby refining segmentation precision.

……

Revised Description in Subsection 3.1

……

With the rapid expansion of deep learning methods based on CNNs, semantic segmentation models have increasingly become the standard in landslide detection (Lu et al., 2023; Zhou et al., 2024). As a fundamental task in computer vision, semantic segmentation assigns a specific class label (e.g., "landslide" or "non-landslide") to each pixel in an image, thereby enabling dense, pixel-level classification (Guo et al., 2018).

Numerous advanced semantic segmentation networks have been proposed and validated for automatic landslide detection, significantly enhancing the efficiency and accuracy of large-scale detection. U-Net is a typical example (Ronneberger et al., 2015), which features a U-shaped architecture. U-Net employs an encoder-decoder structure, where the encoder is similar to conventional CNNs, progressively reducing image resolution and extracting features through convolution and pooling operations; the decoder then restores the image resolution through transposed convolution or upsampling operations (Dong et al., 2022; Nava et al., 2022). Skip connections bridge low-level detail features with deep semantic features, thereby refining segmentation precision.

……

Newly Added References

Guo, Y., Liu, Y., Georgiou, T. and Lew, M.S., 2018. A review of semantic segmentation using deep neural networks. International journal of multimedia information retrieval, 7(2), pp.87-93.doi: 10.1007/s13735-017-0141-z.

Lu, Z., Peng, Y., Li, W., Yu, J., Ge, D., Han, L. and Xiang, W., 2023. An iterative classification and semantic segmentation network for old landslide detection using high-resolution remote sensing images. IEEE Transactions on Geoscience and Remote Sensing, 61, pp.1-13. doi:10.1109/TGRS.2023.3313586.

Zhou, N., Hong, J., Cui, W., Wu, S. and Zhang, Z., 2024. A multiscale attention segment network-based semantic segmentation model for landslide remote sensing images. Remote Sensing, 16(10), p.1712. doi:10.3390/rs16101712.

Comment #20:

19. L285-287: Why is DeepLab preferable to U-Net? Can you elaborate on what ASPP is or does?

Response:

- We sincerely appreciate your valuable comments. Your questions are highly pertinent and have provided us with an opportunity to clarify and further strengthen the content of our manuscript. Below, we provide a point-by-point response and corresponding revisions to address the two concerns you raised.

- *** Regarding the explanation of why DeepLab is preferred over U-Net**
- We selected DeepLab as the primary architecture, primarily due to its superior ability to handle multi-scale objects and delineate smooth boundaries in complex environments—capabilities that are essential for extracting landslides, which are highly variable in morphology and scale.
- While U-Net performs strongly in semantic segmentation owing to its classic encoder–decoder structure and skip connections, DeepLab offers distinct advantages in several critical aspects:
- **Receptive field and multi-scale information capture:** U-Net primarily enlarges the receptive field through pooling operations, which inevitably reduces spatial resolution. Although the decoder partially restores this information via upsampling and skip connections, fine details may still be lost. In contrast, DeepLab leverages dilated (atrous) convolutions to exponentially expand the receptive field without increasing parameter count or reducing feature map resolution. This enables the network to integrate contextual information across broader areas—an essential capability for accurately identifying landslides, which require global environmental cues (e.g., topography, vegetation patterns) for reliable detection.
- **Robustness in complex scenarios:** Landslide regions are typically characterized by substantial complexity, with scales ranging from small shallow failures to large deep-seated slides, and boundaries that are often diffuse and irregular. DeepLab incorporates the unique Atrous Spatial Pyramid Pooling (ASPP) module, specifically designed to address multi-scale challenges by capturing contextual information at multiple scales in parallel. In contrast, U-Net's multi-scale capabilities rely more implicitly on features from different encoder levels, making its multi-scale fusion less explicit and less powerful than that of ASPP.
- **Boundary segmentation accuracy:** DeepLab further enhances boundary delineation by incorporating a dedicated decoder module. After extracting rich semantic features through the encoder (including ASPP), the decoder progressively restores spatial detail. This combined strategy retains strong semantic representation while markedly improving boundary localization accuracy.
- In summary, we do not contend that U-Net is not an excellent model; rather, for the specific task of landslide identification—characterized by highly complex morphologies and multi-scale variability—the DeepLab family (particularly its ASPP structure and dilated convolutions) provides a theoretically more suitable framework. Numerous recent remote sensing segmentation studies (e.g., Chen et al., 2017; Huang et al., 2024) have likewise demonstrated the superior performance of DeepLab in comparably complex geospatial contexts.

- *** Regarding the detailed explanation of what ASPP is and what it does**
- Thank you for pointing this out. Our original explanation of ASPP was indeed overly brief. Atrous Spatial Pyramid Pooling (ASPP) is a core innovation of the DeepLab family; it is designed to enable robust segmentation of objects across multiple scales.
- **Core idea:** ASPP mimics the way humans observe objects: to recognize an object we examine both its fine local details (small scale/near view) and its surrounding context (large scale/distant view). ASPP implements this by applying multiple parallel dilated (atrous) convolution layers with different dilation rates to the same input feature map, thereby sampling context at multiple scales.
- **Structure:** The ASPP module receives the backbone feature map produced by the encoder. It feeds the feature map into four parallel branches: Three dilated convolution branches: each employs a different dilation rate (e.g., rates of 6, 12 and 18). Larger dilation rates yield larger receptive fields and thus capture increasingly broader contextual information. This is analogous to using multiple "magnifying lenses" simultaneously—one focusing on local detail, another on intermediate regions, and a third on the global scene. A global average pooling branch: this branch pools the feature map to a single global vector, which is then upsampled back to the original spatial dimensions. It captures image-level global context and supplies the network with scene-level semantic information.
- The outputs from all parallel branches are concatenated along the channel dimension.
- Finally, a 1×1 convolution is applied to fuse the multi-scale information from all branches and produce the final multi-scale feature representation.
- **Function:** By this mechanism, ASPP renders the network largely size-invariant with respect to target objects. Whether a landslide is small- or large-scale, one or more ASPP branches will capture the contextual cues most relevant for its identification, thereby substantially improving robustness to scale variation and segmentation accuracy.
- We once again thank you for dedicating your time and effort to reviewing our manuscript and for providing these constructive comments. We hope that the above explanations adequately address your concerns, and we have revised the manuscript accordingly in line with your suggestions (for details, please see new **Subsection 3.1**).

Original Description in Subsection 3.1

……

When dealing with complex features in landslide-prone areas, DeepLab is a more suitable choice (Sandric et al., 2024). Built upon deep convolutional neural networks, DeepLab employs dilated convolutions to expand the receptive field and integrates an atrous spatial pyramid pooling (ASPP) module to capture multi-scale contextual information.

……

Revised Description in Subsection 3.1

……

When dealing with complex features in landslide-prone areas, DeepLab is a more suitable choice than U-Net (Sandric et al., 2024). While U-Net excels at preserving fine-grained spatial details through its skip-connections, its ability to capture long-range contextual information is limited by its relatively small receptive field. DeepLab, built upon deep convolutional neural networks, addresses this critical limitation by employing dilated convolutions to exponentially expand the receptive field without sacrificing resolution or increasing parameters substantially.

More importantly, DeepLab integrates an atrous spatial pyramid pooling (ASPP) module, which is key to its superior performance on multi-scale objects like landslides (Chen et al., 2017; Huang et al., 2024). The ASPP module operates in parallel on the same feature map using multiple convolutional branches with different dilation rates (e.g., rates=6, 12, 18). Each branch effectively captures contextual information at a different scale, from fine details to broad, image-level contexts (Niu et al., 2018). All these multi-scale features are then concatenated and fused. This allows the network to simultaneously leverage both local textual cues and global contextual cues, thereby significantly improving recognition accuracy and reducing false positives in geologically complex environments.

……

Newly Added References

Chen, L.C., Papandreou, G., Kokkinos, I., Murphy, K. and Yuille, A.L., 2017. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. IEEE transactions on pattern analysis and machine intelligence, 40(4), pp.834-848.doi:10.1109/TPAMI.2017.2699184.

Huang, J., Song, W., Liu, T., Cui, X., Yan, J. and Wang, X., 2024. Submarine Landslide Identification Based on Improved DeepLabv3 with Spatial and Channel Attention. Remote Sensing, 16(22), p.4205. doi:10.3390/rs16224205.

Niu, Z., Liu, W., Zhao, J. and Jiang, G., 2018. DeepLab-based spatial feature extraction for hyperspectral image classification. IEEE Geoscience and Remote Sensing Letters, 16(2), pp.251-255. doi:10.1109/LGRS.2018.2871507.

Comment #21:

21. L297: What are temporal variation curves?

Response:

- Thank you for raising this question. By "temporal variation curves", we refer to plots that depict how pixel- or region-based feature values (e.g., spectral indices, probabilities, or activation responses) change over time across multi-temporal remote sensing images. Such curves provide an intuitive representation of dynamic processes, such as the progressive development of landslides or vegetation change, by showing temporal trajectories rather than a single snapshot. To avoid ambiguity, we have revised the manuscript. Please see the new **Subsection 3.1** for details.

Original Description in Subsection 3.1

> ……
>
> Wang (2023) demonstrates that 3D CNNs can directly process these 3D tensors. These models capture spatial and temporal features using convolutional kernels while transforming multi-temporal image sequences into change hotspot maps or temporal variation curves as output.
>
> ……

Revised Description in Subsection 3.1

> ……
>
> Wang (2023) demonstrates that 3D CNNs can directly process these 3D tensors. These models capture both spatial and temporal dependencies through three-dimensional convolutional kernels, enabling the direct processing of multi-temporal image sequences. The outputs typically take two complementary forms: change hotspot maps, which highlight regions of significant spatial change across time, and temporal variation curves, which illustrate the evolution of pixel- or region-based feature values throughout the temporal sequence. Together, these representations provide intuitive and complementary tools for characterizing dynamic processes in landslide-prone areas, such as the initiation, progression, and spatial distribution of slope failures.
>
> ……

Comment #22:

22. L298-299: Which studies have used attention mechanisms? This paragraph has no citations, so it is impossible for the reader to refer to work that has been done regarding attention.

Response:

- Thank you for drawing our attention to this omission. We agree that citations are crucial to support our statement. We have now revised the manuscript to include references to key studies that have successfully integrated attention mechanisms with CNNs for multi-temporal remote sensing analysis, particularly in landslide hazard monitoring. (Please see the new **Subsection 3.1**)

Original Description in Subsection 3.1

……

Some studies even have integrated attention mechanisms into conventional CNN architectures to enhance the analysis of multi-temporal remote sensing imagery, thereby enabling the identification of landslide hazard evolution over time.

……

Revised Description in Subsection 3.1

……

Some studies even have integrated attention mechanisms into conventional CNN architectures to enhance the analysis of multi-temporal remote sensing imagery, thereby enabling the identification of landslide hazard evolution over time. For example, Meng et al. (2024) proposed a framework based on CNN and optimized Bidirectional Gated Recurrent Unit (BiGRU) with an attention mechanism, designed to forecast landslide displacement with a step-like curve. Dong et al. (2022) proposed L-Unet which combines multi-scale feature fusion with attention modules to improve landslide segmentation performance, particularly at boundaries.

……

Newly Added References

Dong, Z., An, S., Zhang, J., Yu, J., Li, J. and Xu, D., 2022. L-unet: A landslide extraction model using multi-scale feature fusion and attention mechanism. Remote Sensing, 14(11), p.2552. doi:10.3390/rs14112552.

Meng, S., Shi, Z., Peng, M., Li, G., Zheng, H., Liu, L. and Zhang, L., 2024. Landslide displacement prediction with step-like curve based on convolutional neural network coupled with bi-directional gated recurrent unit optimized by attention mechanism. Engineering Applications of Artificial Intelligence, 133, p.108078. doi:10.1016/j.engappai.2024.108078.

Comment #23:

23. L319: You should mention that one reason RNNs struggle to model long term dependencies is exploding and vanishing gradients, and that LSTMs avoid this via their special gate design.

Response:

- Thank you for your insightful comment! We have revised the paragraph to explicitly state that conventional RNNs suffer from exploding and vanishing gradients, which hinder their ability to capture long-term dependencies. We further clarified that LSTMs overcome this limitation through their gate-based design, which mitigates gradient degradation and enables the modeling of long-range temporal dependencies. Please see the new **Subsection 3.2** for details.

Original Description in Subsection 3.2

……

Due to conventional RNNs struggle to model long-term dependencies and limit their applicability to short-term temporal sequences, long short-term memory networks (LSTM) were developed (Wang et al., 2023b).

LSTM is an enhancement of RNNs, primarily processing long sequence data. Compared to standard RNNs, the hidden layer architecture of LSTM is much more complex. By incorporating memory cells and gating mechanisms, LSTM selectively propagates critical information across multiple time steps, thereby effectively capturing long-range temporal dependencies (Landi et al., 2021; Yu et al., 2019).

……

Revised Description in Subsection 3.2

……

To overcome the vanishing gradient problem inherent in RNNs, LSTMs introduce memory cells and gating mechanisms that selectively retain relevant temporal information (Hochreiter and Schmidhuber, 1997; Landi et al., 2021; Sherstinsky, 2020; Smagulova and James, 2019; Yu et al., 2019). This capability allows them to model the cumulative and delayed responses of slopes to prolonged rainfall or reservoir water level fluctuations.

LSTM models have been widely applied in landslide displacement prediction and early warning. Yang et al. (2019) analyzed the relationships among landslide deformation, rainfall, and reservoir water levels, and found that compared with static models, the LSTM approach more accurately captured the dynamic characteristics of landslides and effectively leveraged historical information. Xu and Niu (2018) used a LSTM model to predict the displacement evolution of the Baijiabao landslide using rainfall and hydrological level data, achieving a higher correlation compared with traditional regression models. In another study focused on shallow landslides, Xiao et al (2022) used a week-ahead LSTM model, which exhibited stable performance and improved prediction accuracy in short-term prediction scenarios. Additionally, Gidon et al (2023) constructed a Bi-LSTM model and achieved a detection accuracy of 93% in the Mawiongrim area. This effectively addresses the limitations of traditional methods and can provide a reliable technical solution for disaster early warning in this area as well as other similar landslide-prone areas.

……

Newly Added References

Gidon, J.S., Borah, J., Sahoo, S., Majumdar, S., Fujita, M., 2023. Bidirectional lstm model for accurate and real-time landslide detection: A case study in mawiongrim, meghalaya, india. IEEE Internet of Things Journal 11, 3792–3800. doi:10.1109/JIOT.2023.3326203.

Xiao, Y., Ju, N., He, C., Xiao, Z., & Ma, Z. (2022). Week-ahead shallow landslide displacement prediction using chaotic models and robust LSTM. Frontiers in Earth Science, 10, 965071. doi:10.3389/feart.2022.965071.

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

Comment #24:

24. L352: Transformers are hugely important in ML right now, so it should warrant more discussion here than the authors have given it, in my opinion. For example, elaborate on what self-attention is, and mention that that transformer-based architectures are state of the art in several areas right now (language, imaging, etc). Are there more works in landslides that have used transformers? I appreciate the author's discussion of its computational limitations; this is key, and it may not be the right choice for all practitioners. You should mention that the main drawback of transformers comes from its quadratic complexity, and that this is a current area of research to alleviate this issue.

29. L513: Similar to my comment about AEs/VAEs, you should move this paragraph to when you first introduce Transformers, rather than re-introducing them here.

Response:

- Thank you for these constructive comments! We fully agree that Transformers represent a major advancement in machine learning and warrant a more thorough and better-structured discussion. Accordingly, we have revised the manuscript in the following ways:
- **Expanded introduction of Transformers:** We elaborated on the self-attention mechanism, explaining how it computes pairwise dependencies across the sequence and assigns dynamic attention weights, thereby enabling the model to capture long-range dependencies more effectively than CNNs or RNNs. We also highlighted that Transformer-based architectures are now state-of-the-art across multiple domains, including natural language processing, computer vision, and multimodal learning.
- **Applications in landslide research:** We supplemented the discussion with recent studies where Transformers have been applied to landslide detection, susceptibility mapping, and spatiotemporal hazard assessment. Although the number of works is still limited, these studies show the growing potential of Transformers in this field.
- **Computational limitations and ongoing research:** We clarified that the main drawback of Transformers lies in their quadratic complexity with respect to sequence length, which leads to high computational and memory costs. As the reviewer suggested, we also emphasized that alleviating this limitation is an active research area, with efficient variants being proposed to

address scalability challenges.

- **Reorganization to avoid redundancy:** The detailed explanation of the Transformer's encoder–decoder structure, self-attention mechanism, and multimodal embedding strategy, which previously appeared in two separate sections, has now been consolidated into the first introduction of Transformers. In later sections, we removed the redundant description and instead included concise cross-references. This restructuring improves readability and avoids unnecessary repetition.

Original Description in Subsection 3.2

3.2 Models for Time Series Analysis in Potential Landslide Identification

……

Transformer was originally designed to handle sequential data in natural language processing, which was first introduced by Vaswani in 2017 (Vaswani et al., 2017). Unlike conventional recurrent and convolutional structures, the Transformer employs a self-attention mechanism to directly model the entire sequence.

Since the Transformer has the ability to adaptively learn latent features and patterns within the data, when it comes to processing landslide time series data, it can automatically tweak the model parameters to accommodate diverse landslide scenarios and temporal data variability (Wang et al., 2024a; Zerveas et al., 2021).

Transformer also can analyze positional relationships across the entire sequence, better capturing complex dependencies in long sequences, making it especially suitable for handling large-scale, long-term sequential datasets.

In contrast, RNN-based models exhibit a relatively simple architecture (Li et al., 2021a; Wang et al., 2020b). Their mechanisms are conceptually intuitive, making them more interpretable (see Fig. 3). On the other hand, Transformers are more complex in structure with numerous parameters, necessitating substantial computational resources during early training to process large-scale data, while being susceptible to overfitting on small datasets. Understanding how the model extracts features and makes decisions is not straightforward from large amounts of landslide data, posing challenges for its interpretability and practical deployment.

……

3.5 Models for Data Fusion in Potential Landslide Identification

……

As previously discussed, Transformer has become a universal architecture for processing sequential and multimodal data, owing to its self-attention mechanism and modular design.

Transformer is also composed of stacked encoders and decoders. However, unlike other architectures, the Transformer architecture introduces the self-attention mechanism (Zhao et al., 2021a), which is a crucial innovation. This enables the Transformer to automatically calculate a weight vector for each position in the input sequence based on the relationship between this position and other positions, so as to represent the importance of this position in the entire sequence. Such a weight vector can be regarded as the "attention distribution" of each position in the input sequence, that is, the model determines which positions in the sequence to focus on. By considering all positions in the input sequence simultaneously, Transformer is able to calculate the correlations between each position and other positions in the sequence in parallel (Esser et al., 2021; Huang and Chen, 2023; Zerveas et al., 2021), rather than processing them step by step like CNNs or RNNs.

Transformer can also convert multimodal data, it transforms them into vector representations via different embedding layers into a unified vector representation through different embedding layers. Subsequently, through the use of the self-attention mechanism and multilayer neural networks, these vectors are fused and feature representations are extracted, enabling the model to process and integrate data from various modalities within the same model framework.

……

Revised Description in Subsection 3.2

3.2 Models for Time Series Analysis in Potential Landslide Identification

……

Transformer, first introduced by Vaswani et al. (2017), was originally designed for natural language processing but has since become a cornerstone architecture in modern machine learning, achieving state-of-the-art performance across diverse domains such as computer vision and multimodal learning.

Unlike conventional recurrent or convolutional models, the Transformer is built upon stacked encoder–decoder layers and relies on a key innovation: the self-attention mechanism. This mechanism enables the model to automatically compute a weight vector (i.e., an attention distribution) for each element in the sequence based on its relevance to all other elements. By evaluating all positions simultaneously (Esser et al., 2021; Huang and Chen, 2023; Zerveas et al., 2021), the Transformer efficiently captures global dependencies across long sequences in parallel, making it more effective than RNNs or CNNs at modeling long-range relationships.

When applied to landslide-related time series data, the Transformer can adaptively learn latent temporal features and patterns, automatically adjusting parameters to accommodate diverse landslide scenarios (Wang et al., 2024a; Zerveas et al., 2021). Moreover, its capability to analyze positional relationships across the entire sequence makes it particularly suitable for processing large-scale, long-term sequential datasets. Beyond time series, the flexibility of the self-attention mechanism also enables Transformers to integrate and fuse multimodal data sources, which will be discussed in Section 3.5.

In contrast, RNN-based models exhibit a relatively simple architecture (Li et al., 2021a; Wang et al., 2020b) and are conceptually intuitive, making them more interpretable (see Fig. 3). Transformers, however, are structurally more complex with numerous parameters, requiring substantial computational resources during training and being susceptible to overfitting on small datasets. Moreover, a key drawback of the standard Transformer is its quadratic computational complexity with respect to sequence length, which becomes prohibitive for very long sequences (Zhuang et al., 2023). This also complicates the interpretation of how the model extracts features and makes decisions from large amounts of landslide data, posing challenges for practical deployment. It is worth noting that mitigating this quadratic complexity is an active research area, with many efficient Transformer variants being developed. Therefore, while powerful, the vanilla Transformer may not be the optimal choice for all practitioners, and its computational demands should be carefully considered.

……

3.5 Models for Data Fusion in Potential Landslide Identification

……

As highlighted in Section 3.2, the Transformer's self-attention mechanism and modular architecture make it a universal framework for processing sequential data and enabling multimodal fusion (Zhao et al., 2021a). In this context, the core advantage of the Transformer lies in its ability to integrate diverse input data (e.g., satellite imagery, GPS time series, and geological maps). It achieves this by employing independent embedding layers to convert each modality into a unified vector representation, which is then fused through the self-attention mechanism. This mechanism computes the interactions and correlations among all elements across different modalities, thereby enabling the model to capture cross-modal dependencies and extract joint feature representations within a unified framework. This capability makes the Transformer particularly suitable for landslide studies (Li et al., 2025).

In addition, hybrid models can be applied to heterogeneous data processing. Their strength lies in integrating the advantages of different models, thereby overcoming the limitations of individual approaches and enabling the effective fusion of multimodal, multiscale, and spatiotemporal features.

When landslide identification requires the simultaneous consideration of spatial features (e.g., imagery and topography) and temporal features (e.g., rainfall and displacement monitoring), CNN-LSTM is a commonly used hybrid model (Gao et al., 2024). This combination leverages the spatial perception capacity of CNNs together with the temporal dependency modeling ability of LSTMs, making it particularly suitable for rainfall-topography coupled landslide prediction. When potential landslides are associated with spatial networks, the GNN-Transformer hybrid is well suited for modeling. This combination enables GNNs to capture the topological relationships among landslide sites, while Transformers facilitate cross-modal learning (Liang et al., 2025; Sun et al., 2025), making it applicable to complex geological environments and large-scale data analysis. Three models can also be integrated. For example, Guo et al. (2024) employed a stacking method that combines 1D-CNN, RNNs, and LSTMs, resulting in a CRNN-LSTM ensemble model.

……

Newly Added References

Gao, B., He, Y., Chen, X., Chen, H., Yang, W. and Zhang, L., 2024. A deep neural network framework for landslide susceptibility mapping by considering time-series rainfall. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 17, pp.5946-5969. doi:10.1109/JSTARS.2024.3370218.

Guo, W., Ye, J., Liu, C., Lv, Y., Zeng, Q. and Huang, X., 2024. An approach for predicting landslide susceptibility and evaluating predisposing factors. International Journal of Applied Earth Observation and Geoinformation, 135, p.104217. doi:10.1016/j.jag.2024.104217.

Li, W., Hsu, C.Y., Wang, S., Gu, Z., Yang, Y., Rogers, B.M. and Liljedahl, A., 2025. A multi-scale vision transformer-based multimodal GeoAI model for mapping Arctic permafrost thaw. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. doi:10.1109/JSTARS.2025.3564310.

Liang, J., Wei, X., Chen, M., Wang, Z. and Liang, J., 2025, April. GNN-Transformer Cooperative Architecture for Trustworthy Graph Contrastive Learning. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 39, No. 18, pp. 18667-18675). doi:10.1609/aaai.v39i18.34054.

Sun, Y., Zhu, D., Wang, Y., Fu, Y. and Tian, Z., 2025. GTC: GNN-transformer co-contrastive learning for self-supervised heterogeneous graph representation. Neural Networks, 181, p.106645. doi:10.1016/j.neunet.2024.106645.

Zhuang, B., Liu, J., Pan, Z., He, H., Weng, Y. and Shen, C., 2023. A survey on efficient training of transformers. arXiv preprint arXiv:2302.01107. doi:10.48550/arXiv.2302.01107.

Comment #25:

27. Section 3.4: Based on the way this section is written, it should be titled “Anomaly Detection” rather than “data cleaning”. These are somewhat similar, but distinct topics. Anomaly detection is about identifying data that is outside the norm of its dataset. Data cleaning is about filling in missing values and ensuring overall data quality. To me, anomaly detection fits more with what the authors discussed, and seems more relevant to landslides (ie, detecting when a landslide is imminent due to some abnormality)

26. L448: To me, it’s a little backwards to introduce AEs here when you’ve already introduced VAEs. Organizationally, it would make more sense to introduce AEs earlier before you talk about VAEs.

25.L445: No citations in this paragraph to back up your claim.

Response:

- Thank you for pointing out this inappropriate organizational issue in our manuscript! We sincerely apologize for this oversight, which undoubtedly affected the logical flow of the section.
- (1) We acknowledge that the content of Section 3.4 focuses more on the identification of abnormal patterns or deviations in time-series and spatial data associated with potential landslide events, rather than on the conventional data cleaning process. Therefore, we have revised the title of this section from "**Models for Data Cleaning in Potential Landslide Identification**" to "**Models for Anomaly Detection in Potential Landslide Identification**", and we have also refined the introductory paragraph to clearly emphasize the role of anomaly detection in identifying early-warning signals of landslides.
- (2) We agree that introducing AEs after VAEs may appear logically inverted, as VAEs are indeed an extension of AEs. Following your suggestion, we have revised the manuscript to introduce AEs first and then proceed to VAEs. Specifically, we now begin with a description of the framework of AEs, including their encoder-decoder architecture, reconstruction mechanism, and application in anomaly detection for landslide monitoring. We then transition to VAEs, highlighting their probabilistic formulation and advantages in handling data uncertainty as an extension of the conventional AE. This restructuring improves the logical flow and helps readers better understand the progression from AEs to VAEs.
- (3) We have added several relevant citations to support the claim that AEs and VAEs are effective in unsupervised anomaly detection. These revisions improve both the rigor and readability of the manuscript. Please see the new **Subsection 3.4** for details.
- Once again, we express our sincere gratitude to you for their valuable time and insightful comments, which have greatly improved the quality of our manuscript!

Original Description in Subsection 3.4

3.4 Models for Data Cleaning in Potential Landslide Identification

In potential landslide identification, data cleaning, particularly anomaly detection, is a critical issue (Deijns et al., 2020; Jiang et al., 2020). It can distinguish between normal fluctuations and true anomalies, identifying early signs such as subtle changes in the mountain's state or abnormal trends in surface displacement, thus enabling more accurate landslide hazard assessment. With the rapid development of deep learning, the applications in data cleaning have become increasingly widespread, enabling models to automatically learn latent data patterns and identify potential anomalies.

AEs and their variational counterparts are highly effective in unsupervised data cleaning. These models autonomously learn normal geomechanical patterns from data and flag deviations, achieving effective hazard identification even when labeled anomaly samples are scarce.

The AE is a typical unsupervised learning model consisting of an encoder and a decoder The encoder compresses the input data into low-dimensional features, and then the decoder reconstructs the input. During the training process, the autoencoder learns the intrinsic features and patterns of normal landslide data, so that for normal data, the reconstruction error is small. When abnormal landslide data is input, due to the difference between its features and the distribution of normal data, the reconstruction error will be large.

When performing anomaly detection, a suitable reconstruction error threshold is set. When the reconstruction error of the test data exceeds this threshold, it can be determined as abnormal data. In the anomaly detection of landslide displacement data monitored by sensors, if the error of the displacement data after being reconstructed by AEs during a certain period is significantly higher than the normal level, it may indicate that there is an abnormal situation of potential landslides during this period.

As previously introduced, VAE is an extension of AE. Compared to conventional autoencoders, VAE introduces randomness into the latent space, making it more effective in handling data uncertainty (Li et al., 2020; Park et al., 2018).

During training, VAEs learn the latent distribution of the data and can generate new samples resembling the training set. When input samples deviate significantly from this learned distribution, the VAE fails to reconstruct them accurately, thereby flagging anomalies through elevated reconstruction errors. For landslide monitoring, if a VAE is trained on imagery of stable slopes, it internalizes stable terrain features. When an image significantly differs from the stable region, the model will produce a high reconstruction error, indicating the presence of anomalous data.

In contrast, AEs are well-suited for univariate anomaly detection, particularly for landslide precursor detection, while VAEs capture latent space distributions and are more effective for multivariate anomaly detection.

GANs can also be utilized in data cleaning (Kang et al., 2024; Xia et al., 2022). In data cleaning, the discriminator is trained to distinguish between generated data and real data. When new test data is input, if the discriminator struggles to determine whether it is real or generated data, the test data may significantly deviate from the distribution of normal data, indicating a potential anomaly. In landslide monitoring, data may be influenced by various factors, GANs demonstrate robustness by filtering out such interference, thereby enhancing data cleaning accuracy (Radoi, 2022).

AnoGAN extends conventional GANs by directly incorporating data cleaning as one of its primary objectives (Lin et al., 2023; Thomine et al., 2023). It introduces an additional encoder during training, which maps input data to the latent space. The difference between this latent vector and the latent vector of normal samples generated by the generator serves as the basis for data cleaning.

RNNs, LSTMs, and GRUs are also effective for identifying anomalous patterns in sequential data (Zhang et al., 2022a). In potential landslide identification, these models process time series inputs to learn normal temporal dynamics and trends. When new data deviates significantly from the normal patterns learned by the model, such deviations can be flagged as anomalies. However, these models are primarily used for time series data, performing data cleaning by predicting future values of the sequence. For instance, if displacement measurements exhibit abrupt deviations while rainfall remains within historical norms, the model detects such discrepancies by comparing observed values with predictions based on learned temporal dependencies.

Revised Description in Subsection 3.4

3.4 Models for Anomaly Detection in Potential Landslide Identification

Anomaly detection refers to identifying patterns or observations that significantly deviate from the expected behavior of a system. In the context of landslides, such anomalies often manifest as early-warning signals, including abnormal displacements, acceleration in movement, or changes in surface deformation patterns. Unlike conventional data cleaning, which focuses on correcting missing or inconsistent records, anomaly detection aims to identify these irregularities to support potential landslide identification (Deijns et al., 2020; Jiang et al., 2020).

Among deep learning approaches, autoencoders (AEs) and their probabilistic extension, variational autoencoders (VAEs), are widely adopted for unsupervised anomaly detection.

An AE consists of an encoder and a decoder that learn to reconstruct input data through a compressed latent representation (Hinton and Salakhutdinov, 2006; Nawaz et al., 2024). During training, AEs can learn the intrinsic features and patterns of normal landslide data, so that for normal data, the reconstruction error is small. For normal inputs, the reconstruction error remains small, whereas for anomalous inputs, which deviate from the learned distribution, the reconstruction error becomes significantly larger.

By setting a suitable threshold for reconstruction error, anomalies can be effectively detected. For example, in the anomaly detection of landslide displacement monitored by sensors, if the reconstruction error of displacement data during a specific period is substantially higher than normal, this may indicate the onset of potential landslide activity (Zhou and Paffenroth, 2017).

Building upon AEs, VAEs introduce stochasticity into the latent space to better model uncertainty and variability in natural data (Li et al., 2020; Park et al., 2018).

During training, VAEs learn the latent distribution of the data and can generate new samples resembling the training set (Kumar et al., 2024). When input samples deviate significantly from this learned distribution, the VAE fails to reconstruct them accurately, thereby flagging anomalies through elevated reconstruction errors. For landslide monitoring, if a VAE is trained on imagery of stable slopes, it internalizes stable terrain features. When an image significantly differs from the stable region, the model will produce a high reconstruction error, indicating the presence of anomalous data.

Consequently, both AEs and their variational counterparts are highly effective in unsupervised data cleaning (Pol et al., 2019; Sakurada and Yairi,2014). These models autonomously learn normal geomechanical patterns from data and flag deviations, achieving effective hazard identification even when labeled anomaly samples are scarce.

In contrast, AEs are well-suited for univariate anomaly detection, particularly for landslide precursor detection, while VAEs capture latent space distributions and are more effective for multivariate anomaly detection.

GANs can also be utilized in anomaly detection (Kang et al., 2024; Xia et al., 2022). The discriminator in a GAN is trained to distinguish between real and generated data. When new test data is input, if the discriminator struggles to determine whether it is real or generated data, the test data may significantly deviate from the distribution of normal data, indicating a potential anomaly. In landslide monitoring, data may be influenced by various factors, GANs demonstrate robustness by filtering out such interference, thereby enhancing data cleaning accuracy (Radoi, 2022).

AnoGAN extends conventional GANs by directly incorporating anomaly detection as one of its primary objectives (Lin et al., 2023; Thomine et al., 2023). It introduces an additional encoder during training, which maps input data to the latent space. The difference between this latent vector and the latent vector of normal samples generated by the generator serves as the basis for data cleaning.

RNNs, LSTMs, and GRUs are also effective for identifying anomalous patterns in sequential data (Zhang et al., 2022a). In potential landslide identification, these models process time series inputs to learn normal temporal dynamics and trends. When new data deviates significantly from the normal patterns learned by the model, such deviations can be flagged as anomalies. However, these models are primarily used for time series data, performing data cleaning by predicting future values of the sequence (Li et al., 2024). For instance, if displacement measurements exhibit abrupt deviations while rainfall remains within historical norms, the model detects such discrepancies by comparing observed values with predictions based on learned temporal dependencies.

Newly Added References

Hinton, G.E. and Salakhutdinov, R.R., 2006. Reducing the dimensionality of data with neural networks. science, 313(5786), pp.504-507. doi:10.1126/science.1127647.

Kumar, P., Priyanka, P., Uday, K.V. and Dutt, V., 2024, February. Predictive Modelling of Himalayan Soil Movement: Addressing Imbalance with Synthetic Variational Autoencoder Data in Kamand Valley. In International Congress on Information and Communication Technology (pp. 137-147). Singapore: Springer Nature Singapore. doi:10.1007/978-981-97-3299-9_11.

Li, Q., Yao, C., Yao, X., Zhou, Z., & Ren, K., 2024, Time Series Prediction of Reservoir Bank Slope Deformation Based on Informer and InSAR: A Case Study of Dawanzi Landslide in the Baihetan Reservoir Area, China. Remote Sensing, 16(15), 2688. doi:10.3390/rs16152688.

Nawaz, A., Khan, S.S. and Ahmad, A., 2024. Ensemble of autoencoders for anomaly detection in biomedical data: A narrative review. IEEE Access, 12, pp.17273-17289.doi:10.1109/ACCESS.2024.3360691.

Pol, A.A., Berger, V., Germain, C., Cerminara, G. and Pierini, M., 2019, December. Anomaly detection with conditional variational autoencoders. In 2019 18th IEEE international conference on machine learning and applications (ICMLA) (pp. 1651-1657). IEEE. doi:10.1109/ICMLA.2019.00270.

Sakurada, M. and Yairi, T., 2014, December. Anomaly detection using autoencoders with nonlinear dimensionality reduction. In Proceedings of the MLSDA 2014 2nd workshop on machine learning for sensory data analysis (pp. 4-11). doi:10.1145/2689746.2689747.

Yin, A., Zheng, F., Tan, J. and Wang, Y., 2020. An improved variational auto-encoder with reverse supervision for the obstacles recognition of UGVs. IEEE Sensors Journal, 21(10), pp.11791-11798. doi: 10.1109/JSEN.2020.3013668.

Zhou, C. and Paffenroth, R.C., 2017, August. Anomaly detection with robust deep autoencoders. In Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining (pp. 665-674). doi:10.1145/3097983.3098052.

Comment #26:

30. The description of the mechanisms of landslides in Section 4 is extensive and appreciated. However, not enough time is spent discussing the main point of the section: the application of DEEP LEARNING to identifying these landslides. Lines 632-638 are a good example of what I would have expected more of. This section needs to do a better job answering the questions: What kinds of model (and what data) are generally used for what applications? Why? And so on.

Response:

- Thank you very much for this valuable and constructive comment! We fully agree that Section 4 previously emphasized the mechanistic aspects of landslides but did not sufficiently elaborate on the application of deep learning models for their identification and prediction.
- In response, we have significantly revised Section 4 to include a more detailed and structured discussion on how different deep learning architectures are applied in landslide studies, what types of data they utilize, and why specific models are more suitable for certain tasks.
- Please see the new **Section 4** for details.

Original Description in Section 4

4.1 Application of Deep Learning in the Identification of Rainfall-induced Landslides

……

Following the development of rainfall threshold models, real-time monitoring of historically rainfall-induced landslides is imperative. First, continuous surveillance enables early detection of subtle deformations and precursory anomalies (Guzzetti et al., 2020; Zhu et al., 2023b), facilitating timely reactivation warnings to mitigate secondary hazards to lives and infrastructure. Second, by continuously monitoring rainfall, soil moisture, and groundwater levels, we can support dynamic recalibration of threshold parameters. This data assimilation enhances model adaptability to evolving hydrogeological conditions, ensuring operational relevance across heterogeneous terrains.

……

4.3 Application of Deep Learning in the Identification of Human Activity-induced Landslides

……

Moreover, the triggers of human activity-induced landslides are not only related to natural conditions but also closely associated with dynamic human activities. Consequently, their analysis necessitates the integration of multimodal and cross-scale data to capture coupled environmental and behavioral drivers (see Fig. 6). In engineering operations such as mining or road construction, factors including proximity to potential landslide zones, excavation depth, and slope angles must be rigorously evaluated through geohazard risk assessments. During excavation phases, geotechnical investigations are imperative to identify weak lithological strata or fracture-dense zones predisposed to instability. Continuous slope stability monitoring requires deploying IoT-enabled sensors to track temporal variations in surface fissure dimensions and subsurface displacement vectors. Monitoring data from these sensors can be integrated into deep learning models for multimodal analytics, enabling dynamic risk prediction and adaptive mitigation planning.

To mitigate misclassification between anthropogenic signatures and natural terrain, integrating multispectral data with topographic elevation data enhances discriminative capacity (Meng et al., 2021; Selamat et al., 2023). For instance, in mountainous regions, DEMs revealing artificially excavated steep slopes combined with fractured geological strata from structural maps provide preliminary evidence of human influence on landslide susceptibility (Lian et al., 2024).

……

Unlike simpler landslide types, identifying composite landslides necessitates multimodal data fusion to holistically assess predisposing conditions (Li, 2025; Yin et al., 2023). It further requires disentangling the nonlinear superposition effects of multiple factors and quantifying their relative contributions to failure initiation. For instance, Dou et al. (2019) analyzed how earthquake intensity and rainfall metrics jointly modulate landslide susceptibility, deriving failure probabilities under varying parameter combinations. In multi factor-induced landslides, earthquakes and rainfall often interact with other factors. For instance, during heavy rainfall, the rate of landslide formation after an earthquake may be higher, possibly driven by the removal of excessively steep slopes, changes in vegetation and groundwater, and alterations in the mechanical properties of the bedrock and weathered layers in the earthquake-induced landslides canopy. This necessitates systematic investigation of multi-hazard coupling effects to quantify emergent risks.

……

With the accumulation of new data and the dynamic variations in multi factor-induced landslides, regular model updates are critical to ensuring identification accuracy and adaptability. Existing studies predominantly apply these methods based on comprehensive historical landslide datasets and employ batch learning theory for identification. However, when new data becomes available, the model must be retrained from scratch. This approach is not only highly inefficient but also fails to account for the connections between newly observed and historical landslides. To address this limitation, incremental learning methods offer a promising solution. These methods enable gradual parameter optimization through new data without retraining the existing model (Huang et al., 2022). Compared to conventional deep learning models, those integrated with incremental learning can more effectively leverage historical landslide data and adaptively learn from newly incorporated data, thereby better accommodating the dynamic nature of landslides.

……

Revised Description in Section 4

4.1 Application of Deep Learning in the Identification of Rainfall-induced Landslides

……

Following the development of rainfall threshold models, real-time monitoring of historically rainfall-induced landslides is imperative. First, continuous surveillance enables early detection of subtle deformations and precursory anomalies (Guzzetti et al., 2020; Zhu et al., 2023b), facilitating timely reactivation warnings to mitigate secondary hazards to lives and infrastructure. Second, by continuously monitoring rainfall, soil moisture, and groundwater levels, we can support dynamic recalibration of threshold parameters. This data assimilation enhances model adaptability to evolving hydrogeological conditions, ensuring operational relevance across heterogeneous terrains.

While the physical mechanisms governing rainfall-induced slope failures have been well studied (Arnone et al., 2011; Xiong et al., 2024), recent advances in deep learning have significantly improved our ability to automatically identify and predict such events using heterogeneous data.

In the context of rainfall-induced landslides, spatiotemporal data (e.g., rainfall intensity, cumulative precipitation, soil moisture, and slope displacement time series) are the primary inputs. Deep learning models are selected according to data characteristics and task objectives. For instance, CNNs are commonly used to extract spatial rainfall–topography features and delineate susceptible zones from remote sensing images (Peng and Wu 2024; Xu et al., 2022; Zhang et al., 2022). The encoder–decoder architecture, such as U-Net, enables pixel-level segmentation of rainfall-triggered landslides (Bhatta et al., 2025), with the inclusion of vegetation or soil moisture layers improving feature discrimination.

When temporal evolution is essential, RNNs and LSTM networks effectively model sequential dependencies between rainfall and slope deformation (Biniyaz et al., 2022; Liu et al., 2025). These models are capable of learning hysteretic responses and time lags between precipitation events and ground displacement, enabling early warning through time-series forecasting.

Furthermore, hybrid models combining CNN and LSTM components have been proposed to jointly learn spatial–temporal correlations (Chen and Fun 2022; Wu et al., 2025). By fusing rainfall distribution maps with displacement monitoring sequences, these architectures provide a more complete understanding of rainfall–landslide coupling mechanisms.

Deep learning also facilitates data-driven rainfall threshold estimation. Instead of relying solely on empirical or physically based thresholds, models such as fully connected neural networks (FNNs) and attention-based transformers can derive adaptive rainfall thresholds from multi-year rainfall–landslide records, capturing regional nonlinearities (Wu et al., 2023).

……

4.3 Application of Deep Learning in the Identification of Human Activity-induced Landslides

……

Moreover, the triggers of human activity-induced landslides are not only related to natural conditions but also closely associated with dynamic human activities (Tian et al., 2025). Consequently, their analysis necessitates the integration of multimodal and cross-scale data to capture coupled environmental and behavioral drivers (see Fig. 6). In engineering operations such as mining or road construction, factors including proximity to potential landslide zones, excavation depth, and slope angles must be rigorously evaluated through geohazard risk assessments. During excavation phases, geotechnical investigations are imperative to identify weak lithological strata or fracture-dense zones predisposed to instability. Continuous slope stability monitoring requires deploying IoT-enabled sensors to track temporal variations in surface fissure dimensions and subsurface displacement vectors. Monitoring data from these sensors can be integrated into deep learning models for multimodal analytics, enabling dynamic risk prediction and adaptive mitigation planning.

For spatial mapping and fissure extraction, CNNs and U-Net-based segmentation models have demonstrated strong capability in identifying artificial slope features from optical or SAR imagery. CNN-based models can capture high-level semantic information on excavation scars, road cuts, and spoil heaps that indicate anthropogenic disturbance. Tao et al. (2024) employed the DRs-UNet model to investigate the use of deep learning for UAV-based crack identification, the developmental patterns of fissures, and the feedback interactions between underground mining progress and corresponding surface conditions. Wu et al. (2021) proposed the PUNet model for detecting and mapping localized rapid subsidence induced by mining activities. Meng et al. (2025) introduced the GF-Former model to achieve precise segmentation of ground fissures in remote sensing imagery.

When surface deformation time series or micro-displacement data from GB-InSAR, InSAR, or IoT sensors are available, RNN-based models are applied to model the temporal evolution of slope deformation (Han et al., 2022; Li et al., 2025). These models are particularly effective in detecting precursory motion trends caused by progressive excavation or loading activities.

To mitigate misclassification between anthropogenic signatures and natural terrain, integrating multispectral data with topographic elevation data enhances discriminative capacity (Meng et al., 2021; Selamat et al., 2023). For instance, in mountainous regions, DEMs revealing artificially excavated steep slopes combined with fractured geological strata from structural maps provide preliminary evidence of human influence on landslide susceptibility (Lian et al., 2024).

……

Unlike simpler landslide types, identifying composite landslides necessitates multimodal data fusion to holistically assess predisposing conditions (Li, 2025; Yin et al., 2023). It further requires disentangling the nonlinear superposition effects of multiple factors and quantifying their relative contributions to failure initiation. For instance, Dou et al. (2019) analyzed how earthquake intensity and rainfall metrics jointly modulate landslide susceptibility, deriving failure probabilities under varying parameter combinations. In multi factor-induced landslides, earthquakes and rainfall often interact with other factors. For instance, during heavy rainfall, the rate of landslide formation after an earthquake may be higher, possibly driven by the removal of excessively steep slopes, changes in vegetation and groundwater, and alterations in the mechanical properties of the bedrock and weathered layers in the earthquake-induced landslides canopy. This necessitates systematic investigation of multi-hazard coupling effects to quantify emergent risks.

In addition to constructing physics-based models that account for multiple factors and quantify their interactions through the solution of governing equations, GNNs can also be employed (Lei et al., 2025). These models are capable of capturing the spatiotemporal dependencies and nonlinear couplings among various triggering factors. By integrating cross-attention mechanisms, their ability to dynamically weight the relationships among stresses induced by rainfall, earthquakes, and human activities can be further enhanced. For example, Ren et al. (2025) employed a Graph Neural Network (GNN) to capture and model the complex spatiotemporal dependencies among multiple monitoring locations during landslide deformation. Zeng et al. (2022) used the graphical representation capability of the GNN model to analyze environmental relationships within a study region, where nodes were defined as geographic units delineated by terrain surface approximations, and edges captured the interactions between node pairs. Zhang et al. (2024) constructed a geographically constrained relational graph based on node features representing environmental similarity and employed a cosine similarity approach to associate landslides with their surrounding geographic environments.

Cross-attention mechanisms can also be integrated into the model to capture spatiotemporal dependencies among contributing factors. For instance, Hu et al. (2025) integrated global landslide feature vectors with local feature maps through a cross-attention mechanism to enhance the discriminative capability between landslides and background geomorphology. Alternatively, gated fusion units may be incorporated to dynamically adjust the weights of multi-modal features (Yang et al., 2024a). For instance, Liu et al. (2022) proposed a gated fusion unit (GFU) module for multimodal remote sensing image semantic classification, enabling early fusion of heterogeneous modality features.

With the accumulation of new data and the dynamic variations in multi factor-induced landslides, regular model updates are critical to ensuring identification accuracy and adaptability. Existing studies predominantly apply these methods based on comprehensive historical landslide datasets and employ batch learning theory for identification. However, when new data becomes available, the model must be retrained from scratch. This approach is not only highly inefficient but also fails to account for the connections between newly observed and historical landslides. To address this limitation, incremental learning methods offer a promising solution. These methods enable gradual parameter optimization through new data without retraining the existing model (Huang et al., 2022). Compared to conventional deep learning models, those integrated with incremental learning can more effectively leverage historical landslide data and adaptively learn from newly incorporated data, thereby better accommodating the dynamic nature of landslides.

……

Newly Added References

Arnone, E., Noto, L.V., Lepore, C. and Bras, R.L., 2011. Physically-based and distributed approach to analyze rainfall-triggered landslides at watershed scale. Geomorphology, 133(3-4), pp.121-131. doi:10.1016/j.geomorph.2011.03.019.

Bhatta, S., Roy, A. and Shahandashti, M., 2025. Land Cover Classification Using U-Net for Calibration of Rainfall-Induced Slope Susceptibility Maps. In International Conference on Transportation and Development 2025 (pp. 439-448). doi:10.1061/9780784486191.039.

Biniyaz, A., Azmoon, B., Sun, Y. and Liu, Z., 2022. Long short-term memory based subsurface drainage control for rainfall-induced landslide prevention. Geosciences, 12(2), p.64. doi:10.3390/geosciences12020064.

Chen, C. and Fan, L., 2022. CNN-LSTM-attention deep learning model for mapping landslide susceptibility in Kerala, India. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 10, pp.25-30. doi:10.5194/isprs-annals-X-3-W1-2022-25-2022.

Han, J., Yang, H., Liu, Y., Lu, Z., Zeng, K. and Jiao, R., 2022. A deep learning application for deformation prediction from ground-based insar. Remote Sensing, 14(20), p.5067. doi:10.3390/rs14205067.

Hu, W., Sun, G., Zeng, X., Tong, B., Wang, Z., Wu, X. and Song, P., 2025. Hierarchical cross attention achieves pixel precise landslide segmentation in submeter optical imagery. Scientific Reports, 15(1), p.21933. doi:10.1038/s41598-025-08695-8.

Li, J., Fan, C., Zhao, K., Zhang, Z. and Duan, P., 2025. Landslide displacement prediction using time series InSAR with combined LSTM and TCN: application to the Xiao Andong landslide, Yunnan Province, China. Natural Hazards, 121(4), pp.3857-3884. doi:10.1007/s11069-024-06937-y.

Liu, Q., Kampffmeyer, M., Jenssen, R. and Salberg, A.B., 2022. Multi-modal land cover mapping of remote sensing images using pyramid attention and gated fusion networks. International Journal of Remote Sensing, 43(9), pp.3509-3535. doi10.1080/01431161.2022.2098078.

Liu, Y., Brezzi, L., Liang, Z., Gabrieli, F., Zhou, Z. and Cola, S., 2025. Image analysis and LSTM methods for forecasting surficial displacements of a landslide triggered by snowfall and rainfall. Landslides, 22(3), pp.619-635. doi:10.1007/s10346-024-02328-3.

Meng, J., Xu, X., Li, P., Zhang, Z., Zhao, W., Ren, J. and Li, Y., 2025. Gf-former: an accurate UAV-based remote sensing image network for high-precision automatic segmentation of ground fissures in mining regions. International Journal of Machine Learning and Cybernetics, pp.1-22. doi:10.1007/s13042-025-02555-7.

Peng, B. and Wu, X., 2024. Optimizing rai.nfall-triggered landslide thresholds for daily landslide hazard warning in the Three Gorges Reservoir area. Natural Hazards and Earth System Sciences, 24(11), pp.3991-4013. doi:10.5194/nhess-24-3991-2024.

Ren X, Liu W, Yang W, Mao L, Li H. Landslide Deformation Uncertainty Quantification Using Conformalized Graph Neural Networks: A Case Study in Sichuan Province, China. IEEE Access. 2025 May 8. doi:10.1109/ACCESS.2025.3568273.

Tao, T., Han, K., Yao, X., Chen, X., Wu, Z., Yao, C., Tian, X., Zhou, Z. and Ren, K., 2024. Identification of ground fissure development in a semi-desert aeolian sand area induced from coal mining: Utilizing UAV images and deep learning techniques. Remote Sensing, 16(6), p.1046. doi:10.3390/rs16061046.

Tian, N., Lan, H., Li, L., Peng, J., Fu, B. and Clague, J.J., 2025. Human activities are intensifying the spatial variation of landslides in the Yellow River Basin. Science Bulletin, 70(2), pp.263-272. doi:10.1016/j.scib.2024.07.007.

Wu, H., Niu, J., Li, Y., Wang, Y. and Qiu, D., 2025. Landslide Susceptibility Prediction Based on a CNN–LSTM–SAM–Attention Hybrid Model. Applied Sciences, 15(13), p.7245. doi:10.3390/app15137245.

Wu, S., Li, X. and Chen, D., 2023, May. A Method of Rainfall-Runoff Prediction Based on Transformer. In Proceedings of the 15th International Conference on Digital Image Processing (pp. 1-6). doi:10.1145/3604078.3604095.

Wu, Z., Wang, T., Wang, Y., Wang, R. and Ge, D., 2021. Deep learning for the detection and phase unwrapping of mining-induced deformation in large-scale interferograms. IEEE Transactions on Geoscience and Remote Sensing, 60, pp.1-18. doi:10.1109/TGRS.2021.3121907.

Xiong, J., Pei, T. and Qiu, T., 2024. A Novel Framework for Spatiotemporal Susceptibility Prediction of Rainfall-Induced Landslides: A Case Study in Western Pennsylvania. Remote Sensing, 16(18), p.3526. doi: 10.3390/rs16183526.

Xu, G., Wang, Y., Wang, L., Soares, L.P. and Grohmann, C.H., 2022. Feature-based constraint deep CNN method for mapping rainfall-induced landslides in remote regions with mountainous terrain: An application to Brazil. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 15, pp.2644-2659. doi: 10.1109/JSTARS.2022.3161383.

Zeng, H., Zhu, Q., Ding, Y., Hu, H., Chen, L., Xie, X., Chen, M. and Yao, Y., 2022. Graph neural networks with

constraints of environmental consistency for landslide susceptibility evaluation. International journal of geographical information science, 36(11), pp.2270-2295. doi:10.1080/13658816.2022.2103819.
Zhang, D., Wei, K., Yao, Y., Yang, J., Zheng, G. and Li, Q., 2022. Capture and prediction of rainfall-induced landslide warning signals using an attention-based temporal convolutional neural network and entropy weight methods. Sensors, 22(16), p.6240. doi: 10.3390/s22166240.
Zhang, Q., He, Y., Zhang, L., Lu, J., Gao, B., Yang, W., Chen, H. and Zhang, Y., 2024. A landslide susceptibility assessment method considering the similarity of geographic environments based on graph neural network. Gondwana Research, 132, pp.323-342. doi:10.1016/j.gr.2024.04.013.

Comment #27:

31. L690: “gated fusion units” are never defined, and I’m not familiar with what they are. Could you define them?

Response:

- Thank you very much for raising this concern! We apologize for the confusion. To clarify, we have presented relevant explanations.
- Although "gated fusion unit" is not a strictly standardized term, it derives from the concept of gated fusion or gated multimodal units, which are commonly employed in fields such as multimodal learning, visual question answering, and video recognition.
- In this study, the term 'gated fusion units' refers to neural modules that integrate heterogeneous features using learnable gating mechanisms, dynamically controlling the contribution of each modality or feature channel. The concept is related to the Gated Multimodal Unit (GMU) proposed by Arevalo et al. (2017).
- To avoid potential misunderstandings, we have included a definition and explanation of the “gated fusion units” in the manuscript.

Original Description in Subsection 4.4

……

In addition to the approach of constructing physics-based models that account for multiple factors, GNNs can be employed. These models represent landslide-prone areas as graph nodes, dynamically updating node states through spatiotemporal edge (Lei et al., 2025). Furthermore, cross-attention mechanisms can be integrated into the model to capture spatiotemporal dependencies among contributing factors. Alternatively, gated fusion units may be incorporated to dynamically adjust the weights of multi-modal features (Yang et al., 2024a).

……

Revised Description in Subsection 4.4

……

In addition to the approach of constructing physics-based models that account for multiple factors, GNNs can be employed. These models represent landslide-prone areas as graph nodes, dynamically updating node states through spatiotemporal edge (Lei et al., 2025). Furthermore, cross-attention mechanisms can be integrated into the model to capture spatiotemporal dependencies among triggering factors. Another noteworthy fusion strategy is the gated fusion unit. Inspired by the gating structures in recurrent neural networks (Arevalo et al., 2017; Kumar et al., 2020; Tsai et al., 2019), this mechanism learns dynamic weights (typically implemented through gating functions such as Sigmoid) to adaptively regulate the information flow of features from different modalities, thereby emphasizing salient features and suppressing noise. Compared with cross-attention, the gated fusion mechanism is generally more lightweight and provides an alternative approach for multimodal feature fusion (Yang et al., 2024a).

……

Newly Added References

Arevalo, J., Solorio, T., Montes-y-Gómez, M. and González, F.A., 2017. Gated multimodal units for information fusion. arXiv preprint arXiv:1702.01992. doi:10.48550/arXiv.1702.01992.

Kumar, A. and Vepa, J., 2020, May. Gated mechanism for attention based multi modal sentiment analysis. In ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 4477-4481). IEEE. doi:10.1109/ICASSP40776.2020.9053012.

Tsai, Y.H.H., Bai, S., Liang, P.P., Kolter, J.Z., Morency, L.P. and Salakhutdinov, R., 2019, July. Multimodal transformer for unaligned multimodal language sequences. In Proceedings of the conference. Association for computational linguistics. Meeting (Vol. 2019, p. 6558). doi:10.18653/v1/p19-1656.

Comment #28:

32. L692: Paragraph may benefit from discussion of “fine-tuning” from the ML literature.

Response:

- Thank you for this excellent suggestion. We agree that introducing the concept of "fine-tuning" provides a more nuanced and technically accurate progression of ideas in this paragraph. In the revised manuscript, we have restructured this section to first discuss the common practice of fine-tuning with new data, then point out its limitations concerning catastrophic forgetting in non-stationary environments like landslides, and finally introduce incremental learning as a more robust solution to these limitations. This change, we believe, significantly strengthens the logical flow and technical depth of our argument.
- Please see the new **Subsection 4.4** for details.

Original Description in Subsection 4.4

……

With the accumulation of new data and the dynamic variations in multi factor-induced landslides, regular model updates are critical to ensuring identification accuracy and adaptability. Existing studies predominantly apply these methods based on comprehensive historical landslide datasets and employ batch learning theory for identification. However, when new data becomes available, the model must be retrained from scratch. This approach is not only highly inefficient but also fails to account for the connections between newly observed and historical landslides. To address this limitation, incremental learning methods offer a promising solution. These methods enable gradual parameter optimization through new data without retraining the existing model (Huang et al., 2022). Compared to conventional deep learning models, those integrated with incremental learning can more effectively leverage historical landslide data and adaptively learn from newly incorporated data, thereby better accommodating the dynamic nature of landslides.

……

Revised Description in Subsection 4.4

……

With the accumulation of new data and the dynamic variations in multi-factor-induced landslides, regular model updates are critical to ensuring identification accuracy and adaptability. Existing studies predominantly apply deep learning methods based on comprehensive historical landslide datasets. However, when new data becomes available, a naive approach is to retrain the model from scratch, which is computationally inefficient and fails to capture the connections between new observations and historical knowledge. A common strategy from the machine learning literature to address this is fine-tuning, where a model pre-trained on a historical dataset is further trained on new data (Süalp et al., 2025). While this avoids full retraining, standard fine-tuning can still lead to catastrophic forgetting of previously learned patterns.

To better accommodate the dynamic nature of landslides, incremental learning methods offer a more advanced and promising solution (Huang et al., 2022; Wang et al., 2024). These methods enable the model to continuously learn from new data streams, gradually optimizing parameters while striving to preserve knowledge from previous tasks. Compared to models that require retraining or basic fine-tuning (Zhao et al., 2024), models integrated with incremental learning can more effectively leverage historical data and adaptively incorporate new information, thereby enhancing long-term adaptability (Zhen et al., 2025).

……

Newly Added References

Süalp, E. and Rezaei, M., 2025. Mitigating Catastrophic Forgetting in Continual Learning through Model Growth. arXiv preprint arXiv:2509.01213. doi:10.48550/arXiv.2509.01213.

Wang, X., Nie, W., Xie, W. and Zhang, Y., 2024. Incremental learning-random forest model-based landslide susceptibility analysis: A case of Ganzhou City, China. Earth Science Informatics, 17(2), pp.1645-1661. doi:10.1007/s12145-024-01229-2.

Zhao, L., Zhang, X., Yan, K., Ding, S. and Huang, W., 2024. Safe: Slow and fast parameter-efficient tuning for continual learning with pre-trained models. Advances in Neural Information Processing Systems, 37, pp.113772-113796.

Zhen, J., Lai, F., Shiau, J.S., Huang, M., Lu, Y. and Lin, J., 2025. An unsupervised incremental learning model to predict geological conditions for earth pressure balance shield tunneling. Journal of Rock Mechanics and Geotechnical Engineering. doi:10.1016/j.jrmge.2024.12.018.

Comment #29:

34. L805: Paragraph may benefit from discussion of Physics-Informed Neural Networks, which may be of interest to scientists in this area.

Response:

- Thank you so much for your constructive and valuable comments! We fully agree that Physics-Informed Neural Networks (PINNs) are highly relevant to the challenges of landslide identification, as they directly address the tension between data-driven feature extraction and the underlying physical processes governing slope stability.
- Your suggestion aligns well with **Subsection 6.3** of our manuscript, where we also emphasize the integration of data-driven models with physical mechanisms. Following your advice, we have accordingly expanded the relevant discussion. Specifically, we now highlight that PINNs embed physical laws into the training objective of neural networks, thereby constraining data-driven learning with domain knowledge. This integration reduces the reliance on large annotated datasets and enhances cross-regional transferability, which is particularly valuable in landslide studies where data are scarce and heterogeneous. Furthermore, we point out that although applications of PINNs to landslide research remain limited, they represent a promising direction for bridging data-driven models with physically grounded mechanisms. (Please see the revised **Subsection 6.3** for details.)

Original Description in Subsection 6.3

……

In the second stage, mechanistic constraints are integrated into the data-driven model to achieve knowledge-data dually driven fusion.

Before model construction, prior knowledge can be derived from external sources, including domain expertise, historical data, and physical principles. Alternatively, mechanistic models may be employed to preprocess raw monitoring data. The outputs of mechanistic models or prior knowledge serve as a foundation for initializing parameters in data-driven models (Cui et al., 2024; Liu et al., 2023a; Ma and Mei, 2025). This is because, in data-driven models, the selection of initial parameter values significantly impacts on both the training process and final model performance. Incorporating prior knowledge helps define more reasonable initial parameter ranges, enabling the model to converge toward near-optimal solutions earlier in the training phase.

Knowledge embedding involves translating landslide physics into model constraints to guide the training and optimization of data-driven models (Dahal and Lombardo, 2025; Liu et al., 2024). At the architectural level, layers derived from physical equations can be structurally integrated into the network design. These physical equations can even be directly encoded as network layers, forming differentiable physics-informed computational modules. Differentiability is essential to ensure that these physics-based layers function as effective computational modules within the network. This requirement stems from the fact that training relies on optimization algorithms, which adjust model parameters by computing gradients of the loss function with respect to those parameters. Only differentiable physics-encoded layers allow gradient computation during backpropagation, enabling the model to learn parameters consistent with physical laws. At the loss function level, physical equations can be directly embedded into the neural network's loss function to enforce predictions that adhere to physical principles. As the model seeks to minimize the loss function, it iteratively adjusts its parameters to align predictions with the constraints imposed by these physical equations.

……

Revised Description in Subsection 6.3

……

In the second stage, mechanistic constraints are integrated into the data-driven model to achieve knowledge-data dually driven fusion.

Prior knowledge can be derived from external sources, including domain expertise, historical records, and physical principles, or mechanistic models can be employed to preprocess raw monitoring data. These outputs serve as a foundation for initializing parameters in data-driven models, which is crucial because the choice of initial values substantially affects both training efficiency and final performance (Cui et al., 2024; Liu et al., 2023a; Ma and Mei, 2025). Beyond initialization, knowledge embedding involves translating landslide physics into model constraints to guide learning and optimization (Dahal and Lombardo, 2025; Liu et al., 2024). At the architectural level, physical equations can be structurally encoded as differentiable network layers, enabling gradient-based optimization. At the loss function level, physical constraints can be directly incorporated into the training objective, ensuring that predictions remain consistent with established principles.

A representative example of this paradigm is the Physics-Informed Neural Network (PINN) framework (Raissi et al., 2019). PINNs embed governing equations (such as partial differential equations describing slope hydrology or stress-strain processes) into the neural network training objective, thereby constraining the learning process with domain knowledge. This approach not only reduces dependence on large annotated datasets but also enhances interpretability and cross-regional transferability (Karniadakis et al., 2021). Although applications of PINNs in landslide research remain limited (Moeineddin et al., 2023), they provide a promising avenue for bridging purely data-driven approaches with physically grounded mechanisms (Wu et al., 2022).

……

Newly Added References

Karniadakis, G.E., Kevrekidis, I.G., Lu, L., Perdikaris, P., Wang, S. and Yang, L., 2021. Physics-informed machine learning. Nature Reviews Physics, 3(6), pp.422-440. doi:10.1038/s42254-021-00314-5.

Moeineddin, A., Seguí, C., Dueber, S. and Fuentes, R., 2023. Physics-informed neural networks applied to catastrophic creeping landslides. Landslides, 20(9), pp.1853-1863. doi:10.1007/s10346-023-02072-0.

Raissi, M., Perdikaris, P. and Karniadakis, G.E., 2019. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. Journal of Computational physics, 378, pp.686-707. doi:10.1016/j.jcp.2018.10.045.

Wu, Y., Shao, K., Piccialli, F. and Mei, G., 2022. Numerical modeling of the propagation process of landslide surge using physics-informed deep learning. Advanced Modeling and Simulation in Engineering Sciences, 9(1), p.14. doi:10.1186/s40323-022-00228-6.

Comment #30:

35. L833-841: Perhaps I am misunderstanding, but this information is already conveyed in Section 4 and Figure 7.

Response:

- Thank you for your careful reading and insightful observation! You are correct that the description of the three stages of landslides (early, middle, and late) in lines 833–841 was redundant, as it overlapped with the content already presented in Section 4 and Figure 7. This repetition was indeed unnecessary.
- In accordance with your suggestion, we have made the following revisions to the relevant section:
- (1) We have completely removed the repetitive paragraph in lines 833–841. We recognized that Section 4 already provides a more detailed, case-based discussion of these stages, making

repetition at this point unnecessary.

- (2) After the deletion, we rewrote the transition sentence to directly guide readers back to the key findings in Section 4, thereby maintaining logical flow while avoiding redundancy.
- Several revisions are listed as follows. All related revisions are made and marked in the revised manuscript. Please see the revised **Subsection 5.3.2** for all details. Thanks.

Original Description in Subsection 5.3.2

……

The inducing factors of landslides are not only extremely complex in spatial distribution but also highly dynamic in terms of time (Gao et al., 2023). This variability makes the research process of the landslide mechanism more difficult.

From the perspective of temporal dynamics, landslide formation is not instantaneous but evolves through prolonged stages. From initial deformation to eventual collapse, dynamic changes persist throughout the process, with distinct mechanisms governing each phase.

The early stage of a landslide is typically characterized by minor surface deformations or cracks, many of which remain imperceptible. The absence of conspicuous surface indicators results in the frequent omission of initial deformations, thereby heightening instability risks in later phases.

During the intermediate stage, accelerated deformation and pronounced surface fracturing emerge. At this stage, landslide dynamics grow increasingly complex, influenced by competing mechanical mechanisms. The evolving stress and strain fields complicate precise quantification of failure magnitude and velocity.

The terminal stage involves abrupt destabilization and catastrophic collapse, resulting in extensive surface disruption and mass displacement. Nonlinear dynamics dominate this phase, where rapid progression severely limits the feasibility of timely mitigation efforts.

Since the numerical simulation of long-term creep requires a long-time step, while the dynamic process of short-term abrupt changes requires a time resolution in the microsecond level, it is difficult to establish a unified model for these two situations. This will further intensify the conflict of time scales.

In terms of spatial heterogeneity, the influence scope of landslides usually involves geological structures ranging from the microscopic structure of geotechnical particles to the regional scale. Moreover, there are differences in the stratum structure, slope morphology, vegetation coverage, water content, which makes the effects of the same inducing factor vary in different regions. For example, in loose soil layers, heavy rainfall may lead to shallow landslides, while on rocky slopes with well-developed joints, earthquakes or water level fluctuations may trigger deep-seated landslides.

……

Revised Description in Subsection 5.3.2

……

The inducing factors of landslides are not only extremely complex in spatial distribution but also highly dynamic in terms of time (Gao et al., 2023). This variability makes the research process of the landslide mechanism more difficult.

From the perspective of temporal dynamics, landslide formation is not instantaneous but evolves through prolonged stages, each governed by distinct mechanisms (as detailed in Section 4 and Figure 7). This dynamic progression across different timescales creates a fundamental modeling challenge: since the numerical simulation of long-term creep requires a long time step, while the dynamic process of short-term abrupt changes requires a time resolution in the microsecond level, it is difficult to establish a unified model for these two situations. This will further intensify the conflict of time scales.

In terms of spatial heterogeneity, the influence scope of landslides usually involves geological structures ranging from the microscopic structure of geotechnical particles to the regional scale. Moreover, there are differences in the stratum structure, slope morphology, vegetation coverage, water content, which makes the effects of the same inducing factor vary in different regions. For example, in loose soil layers, heavy rainfall may lead to shallow landslides, while on rocky slopes with well-developed joints, earthquakes or water level fluctuations may trigger deep-seated landslides.

……

Comment #31:

36. Figures 9 and 10 seem redundant to me. It is not clear why you need both instead of just one or the other. Moreover, Figure 10(d) seems completely unnecessary.

Response:

- Thank you very much for your constructive comment! We carefully reviewed Figures 9 and 10 and agree that there is a degree of redundancy between them. Upon re-examination, we found that Figure 10 mainly served as a visual integration of concepts that had already been illustrated in earlier figures:
- (1) **Figure 10(a)** corresponds to the real-time monitoring data source that was already presented in Figure 1.
- (2) **Figure 10(b)** represents the stages of landslide evolution, which had been clearly illustrated in Figure 7.
- (3) **Figure 10(c)** conceptually overlaps with the workflow of the knowledge-data dually driven

paradigm shown in Figure 9.

- (4) As you correctly pointed out, **Figure 10(d)** is indeed redundant and not essential to the overall presentation.
- Therefore, in the revised manuscript, we have deleted Figure 10 to avoid duplication and enhance the clarity and conciseness of the paper. We have also adjusted the corresponding text to ensure a smooth logical transition from the previous figures to Figure 9, which now serves as the central illustration of our proposed framework.
- We sincerely appreciate this valuable suggestion, which helped us improve the structure and readability of the manuscript.

Original Description in Subsection 6.3

……

The knowledge-data dually driven paradigm, operating through an iterative "theory-guided data assimilation and data informed theoretical refinement" mechanism, has advanced potential landslide identification from empirical reliance to scientifically quantifiable methodologies.

Furthermore, the spatial analysis capabilities of geographic information system (GIS) were integrated into the practical identification workflow, enabling the study area to be partitioned into distinct landslide risk categories. This risk stratification considers the combined influence of region-specific factors, ensuring scientifically robust and practically viable classifications.

……

Revised Description in Subsection 6.3

……

The knowledge-data dually driven paradigm, operating through an iterative "theory-guided data assimilation and data informed theoretical refinement" mechanism, has advanced potential landslide identification from empirical reliance to scientifically quantifiable methodologies. **T**he overall workflow of this knowledge-data dually driven paradigm for potential landslide identification is conceptually summarized in Figure 9.

Furthermore, the spatial analysis capabilities of geographic information system (GIS) were integrated into the practical identification workflow, enabling the study area to be partitioned into distinct landslide risk categories. This risk stratification considers the combined influence of region-specific factors, ensuring scientifically robust and practically viable classifications.

……

- With our clarifications and revisions, we hope that we have addressed your concerns. Thank you so much for your kind consideration!
- Have a nice day!
- Pan Jiang & Zhengjing Ma & Gang Mei