



Review

# Deep Learning for Exploring Landslides with Remote Sensing and Geo-Environmental Data: Frameworks, Progress, Challenges, and Opportunities

Qi Zhang and Teng Wang \*

School of Earth and Space Sciences, Peking University, Beijing 100871, China; ethan.qi.zhang@outlook.com

\* Correspondence: wang.teng@pku.edu.cn

**Abstract:** This article offers a comprehensive AI-centric review of deep learning in exploring landslides with remote-sensing techniques, breaking new ground beyond traditional methodologies. We categorize deep learning tasks into five key frameworks—classification, detection, segmentation, sequence, and the hybrid framework—and analyze their specific applications in landslide-related tasks. Following the presented frameworks, we review state-of-the-art studies and provide clear insights into the powerful capability of deep learning models for landslide detection, mapping, susceptibility mapping, and displacement prediction. We then discuss current challenges and future research directions, emphasizing areas like model generalizability and advanced network architectures. Aimed at serving both newcomers and experts on remote sensing and engineering geology, this review highlights the potential of deep learning in advancing landslide risk management and preservation.

**Keywords:** deep learning; remote sensing; landslide detection; landslide mapping; landslide susceptibility mapping; landslide displacement prediction



**Citation:** Zhang, Q.; Wang, T. Deep Learning for Exploring Landslides with Remote Sensing and Geo-Environmental Data: Frameworks, Progress, Challenges, and Opportunities. *Remote Sens.* **2024**, *16*, 1344. <https://doi.org/10.3390/rs16081344>

Academic Editors: Ivan Gratchev, Dong-Hyun Kim and Sung-Yup Ohn

Received: 25 February 2024

Revised: 7 April 2024

Accepted: 9 April 2024

Published: 11 April 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Landslides, a critical global geohazard, present substantial risks to both local communities and natural environments [1]. Triggered by a combination of natural phenomena like heavy rainfall, seismic activities, and human-induced factors such as deforestation and construction, they lead to significant soil and rock displacement [2]. Their unpredictable and widespread nature makes landslides a concern worldwide, with severe consequences including loss of life, infrastructural damage, and lasting environmental harm [3]. This situation underscores the urgent need for advanced detecting, monitoring, prediction and management methods to effectively tackle the complex challenges associated with landslides.

Traditional methods in landslide analysis, including empirical, statistical, and physical modeling, have provided foundational insights [4]. However, these approaches have notable limitations. Empirical methods, dependent on historical data, often fall short in predicting unexpected landslide events, particularly under changing climatic conditions [5,6]. Statistical models are useful for general trend analysis but face challenges in addressing the nonlinear aspects of landslide triggers [7,8]. Physical models, while informative, tend to oversimplify the complex real-world scenarios, which can hinder the accuracy of prediction [9,10]. Moreover, these traditional methods are limited in their capacity to process extensive datasets and adapt to the dynamic, changing nature of environmental conditions [11]. With the increasing frequency and severity of landslides, exacerbated by climate change and human activities, there is a growing necessity for more sophisticated, adapted, and precise predictive tools.

The recent advancements in remote sensing technologies have significantly impacted landslide analysis [12,13]. The availability of high-resolution satellite imagery allows for

the detailed observation of Earth's surface changes over time, which is crucial in detecting early signs of landslides. Synthetic Aperture Radar (SAR) data facilitate this process with their ability to assess surface textures and changes, regardless of weather conditions, providing crucial information about slope stability and erosion [14]. Synthetic Aperture Radar Interferometry (InSAR), with its ability to measure millimeter-level surface movements, significantly contributes to landslide analysis by providing precise measurements of ground deformation and early detection of potential landslide areas [15]. Additionally, multispectral imagery, which collects data across various wavelengths, is useful for analyzing soil composition and moisture content, both critical factors in determining landslide susceptibility [16]. However, traditional methods have inherent drawbacks in harnessing the full potential of these data sources. Their limited ability to process and interpret the vast and complex datasets generated by modern remote sensing technologies highlights the need for more sophisticated analytical tools [17].

Since 2010s, deep learning, a subset of artificial intelligence, is revolutionizing the landslide analysis [18]. Its ability to process and analyze vast amounts of complex, multi-dimensional data sets it apart from traditional methods. Deep learning algorithms can identify intricate patterns in data that are often imperceptible to human analysts or conventional analytical techniques [19]. This capability extends to various types of landslides, such as rockfalls, debris slides, or complex movements, as long as these activities are adequately represented in the observation data, which includes analyzing satellite imagery [20,21], weather patterns [22,23], and geological data to predict potential landslide occurrences [24,25]. Unlike traditional models, deep learning systems can continuously learn and improve their predictive capabilities, adapting to new data and changing environmental conditions. This ability is pivotal for applications such as landslide detection, mapping, susceptibility analysis, and displacement prediction. Comparative studies have consistently demonstrated the superiority of deep learning over traditional models, making it an indispensable tool in modern landslide risk assessment and mitigation strategies [26–28]. In this article, we explore the main frameworks of deep learning applied in landslide studies with remote sensing data, specifically in the tasks of landslide detection, mapping, susceptibility analysis, and displacement prediction. Through these case studies, we aim to highlight the potential of deep learning in landslide risk management, and indicate both challenges and opportunities for future research.

The key contributions of this review include the following:

- (1) We provide a detailed overview of deep learning, tracing its development from early concepts to the latest advancements. This detailed overview establishes a robust foundation in deep learning principles, crucial for understanding its applications in landslide studies with remote sensing.
- (2) We categorize deep learning tasks into five frameworks—classification, detection, segmentation, sequence, and hybrid. This structure not only provides a clear understanding of the application of these methods in earth and environment studies but also offers a novel perspective for readers, especially beneficial for remote-sensing experts who are, however, new to AI.
- (3) Our review stands out by the AI-centric approach in examining deep learning applications to landslides. Instead of a general task-based analysis, we scrutinize how specific deep learning frameworks are adeptly applied to various landslide-related tasks. This focused perspective provides insights into which frameworks are more suitable for overcoming particular challenges in studying landslides.
- (4) We discuss current challenges and highlight potential future research directions, contributing to the ongoing evolution of applying deep learning to landslides.

This review is organized as follows: Section 2 provides an in-depth overview of deep learning methodologies. Section 3 summarizes five prevalent deep learning frameworks used in various fields. Section 4 reviews the latest studies on deep learning approaches for studying landslides with remote sensing, illustrated corresponding to these frameworks. Section 5 discusses the challenges and potential future directions in the field, exploring

both the limitations and opportunities of current methods. Finally, Section 6 concludes the paper, summarizing the main findings and insights.

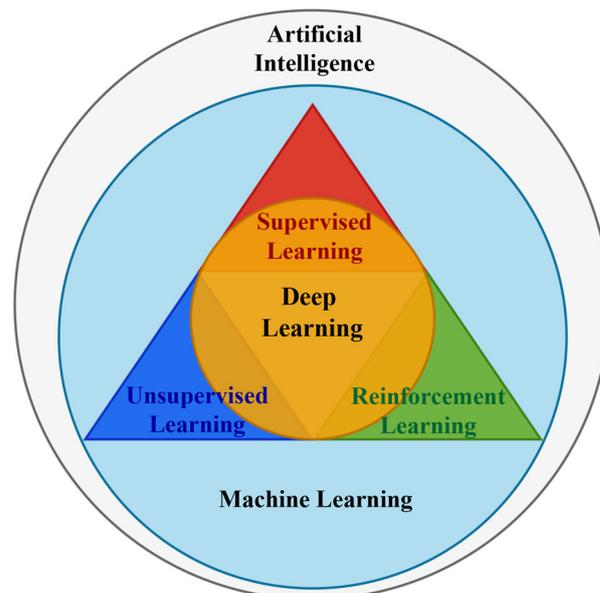
## 2. Deep Learning: Methods, Models, Loss, Evaluation Metrics, Architectural Modules, and Implementing Strategies

### 2.1. Methods

#### 2.1.1. Framework of AI, ML, DL, and Learning Paradigms

In the evolution of artificial intelligence (AI), understanding the relationship between its key components—machine learning (ML), deep learning (DL), and fundamental learning paradigms—is crucial. AI, the overarching concept, encompasses the development of intelligent systems that mimic tasks typically requiring human intelligence. Within AI lies ML, which focuses on algorithms that learn from data to make predictions or decisions. Deep learning, a specialized branch of ML, employs deep neural networks for analyzing and learning from vast volumes of unstructured data, representing a more advanced stage of machine learning [19].

Crucially, both ML and DL encompass three key learning paradigms: supervised, unsupervised, and reinforcement learning, as shown in Figure 1. Each paradigm offers a distinct approach for training algorithms to analyze and learn from data. Therefore, while ML and DL primarily concentrate on the evolution of algorithms and data processing techniques, these learning paradigms provide the specific frameworks within which these technologies function and develop.



**Figure 1.** Hierarchy of AI methods and learning paradigms.

#### 2.1.2. Supervised Learning

Supervised learning, the most prevalent paradigm in ML, focuses on learning a latent function based on labeled input–output pairs. It infers a function from labeled training data, aiming to predict outcomes for unforeseen data. The typical supervised learning algorithm is represented as

$$y = f(x; \theta) \quad (1)$$

where  $y$  and  $x$  represent the output and input,  $\theta$  denotes the parameters of the model, and  $f$  is the learned function. This paradigm is instrumental in fields like image and speech recognition, where the desired output is known, and the model is trained to emulate this mapping. In ML and DL, techniques such as linear regression, support vector machines (SVM), and convolutional neural networks (CNN) are classic examples. The strength of

supervised learning is its effectiveness for specific, well-defined problems, but its reliance on extensive labeled datasets is a considerable limitation.

### 2.1.3. Unsupervised Learning

Different from supervised learning, unsupervised learning aims to identify patterns in data without pre-assigned labels. Its primary goal is to model the underlying structure or distribution of the data to gain deeper insights. Techniques in unsupervised learning vary widely. For example, clustering and dimensionality reduction methods like K-means and principal component analysis (PCA) are commonly used in ML for exploratory data analysis. In DL, approaches such as autoencoders are employed for more complex tasks like feature learning. Unsupervised learning is invaluable for discovering latent patterns in datasets but also faces challenges due to the subjective nature of its results and the difficulty in validating models without definitive labels.

### 2.1.4. Reinforcement Learning

Reinforcement learning is devised for training models to make decisions through trial-and-error interactions with an environment. This paradigm involves an agent learning to achieve a goal in a complex, uncertain environment by performing actions and observing the actions to verify their effectiveness. The learning process is guided by rewards and punishments, akin to how humans learn from real-world interactions [19]. Reinforcement learning is particularly effective in complex decision-making environments, such as game-play, autonomous vehicles, and resource management, but it has limitations in balancing the trade-off between exploring new possibilities and exploiting known strategies, especially in scenarios with sparse or delayed rewards. The comparison of three learning paradigms is shown in Table 1.

**Table 1.** Learning paradigm summarization.

Model	Key Characteristic	Advantages	Limitations	Application
Supervised Learning	Requires labeled datasets and direct feedback on model predictions	Effective in specific and well-defined tasks	Depend on labeled data, may not generalize well beyond training data	Image Classification, regression, detection, segmentation
Unsupervised Learning	No need for labeled data, focuses on data exploration	Discovering unknown patterns in data, works with unlabeled datasets	Less accurate outcomes, hard to evaluate	Clustering, dimensionality reduction, anomaly detection
Reinforcement Learning	Trial-and-error search, delayed reward, interaction with environments	Effective in complex and dynamic environments	Requires a lot of data and computational resources, sensitive to reward structure	Game playing, autonomous vehicles, robotics

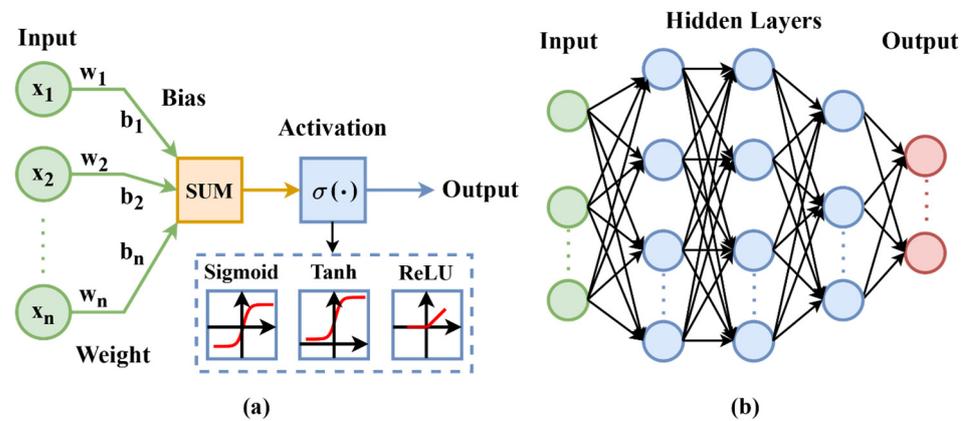
## 2.2. Models

### 2.2.1. Introduction to Neural Network

A neural network [29], pivotal in machine learning, draws inspiration from the complex biological neural networks in animal brains. At its most basic, a neural network, or Artificial Neural Network (ANN), consists of interconnected nodes or units that mirror the neurons in the brain. These are the simplest form of neural networks and are adapted for processing and analyzing complex data structures, as shown in Figure 2a. The core operation of an ANN is captured in the formula

$$y = f\left(\sum_{i=1}^n w_i x_i + b\right) \quad (2)$$

where  $y$  represents the output,  $x_i$  signifies the input features,  $w_i$  refers to the weights assigned to these features,  $b$  is a bias term, and  $f$  is the activation function. This formula is akin to how neurons in the brain process inputs to produce an output.



**Figure 2.** Architectural diagrams of (a) neural network and (b) multi-layer perceptron.

ANNs typically consist of a single layer that directly connects the input to the output, which can limit their ability to handle highly complex tasks. To address this limitation, the field of neural network research has evolved to develop more complex structures such as the Multilayer Perceptron (MLP). An MLP expands upon the basic ANN structure by introducing one or more hidden layers between the input and output layers, allowing the network to learn more complex representations. Each layer in an MLP consists of nodes that function similarly to the basic ANN, but with the added capability to perform more sophisticated computations due to the deeper architecture, as illustrated in Figure 2b. Its architecture can be expressed as

$$y = f^{(L)} \left( \dots f^{(2)} \left( f^{(1)} \left( x^{(1)} W^{(1)} + b^{(1)} \right) W^{(2)} + b^{(2)} \right) \dots W^{(L)} + b^{(L)} \right) \quad (3)$$

where  $L$  denotes the number of layers,  $x^{(1)}$  is the input,  $W^{(1)}$  and  $b^{(L)}$  represent the weights and biases at layer  $L$ , and  $f^{(L)}$  is the activation function for that layer. MLPs demonstrate the capability to capture and represent complex, non-linear data patterns, marking a significant advancement in neural network development.

This formula is akin to how neurons in the brain process inputs to produce an output. Backpropagation, a key step in training neural networks, involves iteratively adjusting the weights of neural networks based on the computed gradients of the loss function. This process allows the network to learn and improve its performance with training. However, during the backpropagation process, the gradients of the networks may become extremely small, hindering the weight updates—particularly problematic in deeper networks, namely the “vanishing gradient problem”.

Activation functions are essential in neural networks, which introduces non-linearity and enables the network to represent complex data patterns beyond simple linear relationships. Widely used activation functions include the following.

**Sigmoid:**  $\sigma(x) = \frac{1}{1+e^{-x}}$ . It is suitable for binary classification, yet susceptible to the vanishing gradient problem, especially in deeper networks.

**ReLU (Rectified Linear Unit):**  $\text{ReLU}(x) = \max(0, x)$ . It is known for accelerating convergence in deep networks and mitigating the vanishing gradient problem to some extent. However, it introduces the “dying ReLU problem”, where neurons can become inactive during training, leading to a complete loss of learning capability for those neurons.

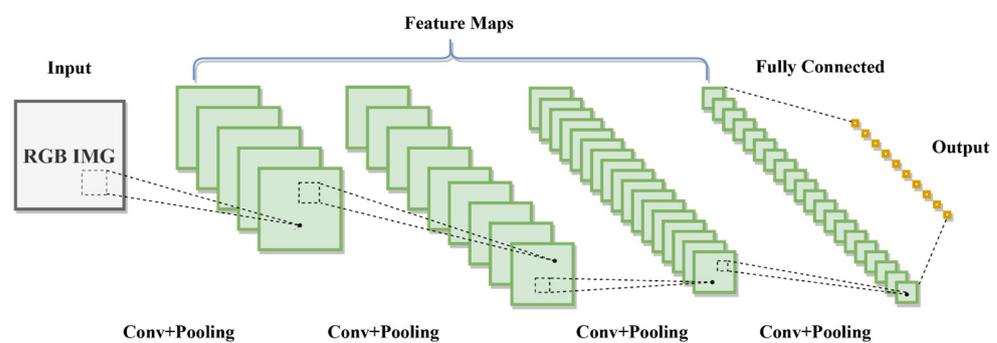
**Hyperbolic Tangent (tanh):**  $\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$ . It projects outputs between  $-1$  and  $1$ , but is still subject to vanishing gradients.

### 2.2.2. Vision Models for Spatial Learning

Spatial learning is focused on understanding the layout, context, and interrelationships within visual data, crucial for tasks like image recognition, object detection, and scene comprehension. CNNs have revolutionized spatial learning, significantly altering the way we interpret and analyze spatial patterns in images [30]. Their architecture consists of convolutional layers, followed by pooling and fully connected layers, enabling the efficient processing of grid-like data and extraction of vital features for various visual tasks, as shown in Figure 3. The convolution operation is defined as

$$(S * K)(i,j) = \sum_m \sum_n S(m,n)K(i-m,j-n) \quad (4)$$

where  $S$  represents the input image,  $K$  denotes the filters, and  $i, j$  are the coordinates in the output feature map. This process involves sliding the filter kernel  $K$  over the input signal  $S$ , multiplying corresponding elements, and summing these products to generate new values in the output feature map.



**Figure 3.** Convolutional neural network (CNN) architecture.

In the spatial feature learning, three tasks are predominant: image classification, object detection, and image segmentation. The evolution of these tasks reflects a parallel and interconnected development, influencing and progressing the others. Image classification, an early focus in computer vision, has been fundamental in neural networks' advancement for image analysis. Initial models like LeNet [31] propelled the field, setting a benchmark for neural networks in image classification. This task established the basis for pattern recognition and feature extraction, influencing subsequent complex tasks. ResNet [32], a groundbreaking work, proposed skip connection and residual block, which enables the possibility of building deeper and more powerful networks. Advancements in image classification have also benefited other spatial learning tasks, contributing to the overall growth of the field.

Concurrently, object detection began to emerge as another significant area of focus. Although it built on the principles established in image classification, such as recognizing patterns and features, object detection introduced an additional challenge, i.e., locating and identifying multiple objects within a single image. This task necessitates a deeper understanding of spatial relationships, leading to innovative networks like the R-CNN [33–35] and the Yolo series [36–38]. Image segmentation, though seemingly more advanced, also developed alongside classification and detection. This task involves dividing an image into pixel-level segments, requiring detailed analysis and dense prediction. Representative networks like Fully Convolutional Networks (FCNs) [39] and U-Net [40], which introduced an encoder–decoder framework, paving a solid foundation in this field. The development of these segmentation networks was influenced by the ongoing progress in both image classification and object detection, showcasing the interdependency and parallel growth among these vision tasks.

Vision models are instrumental in automating the analysis of remote sensing imagery to detect, locate and map the key features of potential landslides. For instance, image

classification models, proficient in identifying distinctive patterns associated with unstable terrain, play an important role in comprehensive landslide feature analysis. Object detection models, with their capability to swiftly pinpoint potential landslide features within large-scale satellite images, offer invaluable advantages in enhancing the efficiency of locating potential landslides. This rapid detection not only streamlines the mapping process but also ensures a focused assessment of the spatial distribution to potential landslides. Moreover, image segmentation models provide a detailed approach by accurately delineating the boundaries of landslide-affected areas. By outlining distinct regions within the imagery, these models enable quantitative measurements and change detection analysis, benefiting a nuanced understanding of the landslide characteristics.

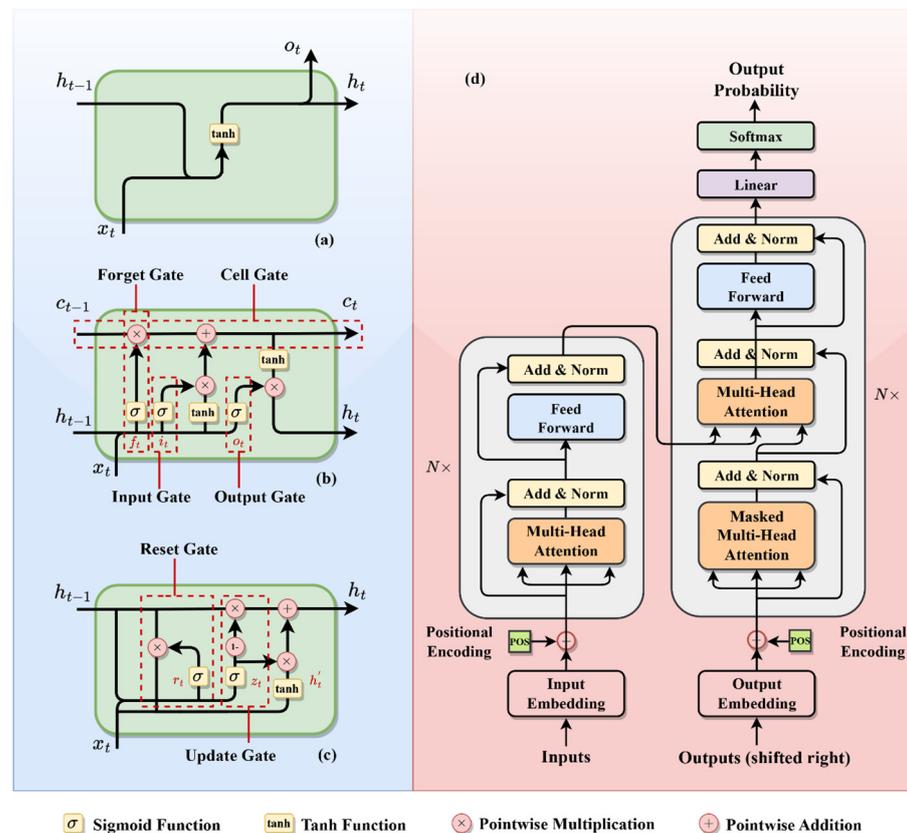
### 2.2.3. Sequence Models for Temporal Learning

Sequence data analysis has long been a crucial aspect of machine learning, with applications ranging from natural language processing and financial forecasting to speech recognition. Over time, the field of sequence modeling has witnessed significant advancements in unraveling the complexities of temporal data.

The initial breakthrough was the introduction of the Recurrent Neural Network (RNN) [41]. RNNs brought in the concept of self-loops, enabling the retention and processing of information over time steps, as shown in Figure 4a. They are expressed by the formula

$$h_t = \sigma(W_{hh}h_{t-1} + W_{xh}x_t + b_h) \tag{5}$$

where  $h_t$  is the hidden state at time  $t$ ,  $x_t$  is the input at time  $t$ ,  $W_{hh}$  and  $W_{xh}$  are weight matrices,  $b_h$  is the bias, and  $\sigma$  is the activation function. While RNNs are effective at capturing short-term dependencies, they struggle with long-term sequences due to the similar vanishing gradient problem as in the spatial learning, making it hard to retain information over long periods.



**Figure 4.** Architectural diagrams of sequence models: (a) RNN; (b) LSTM; (c) GRU; (d) transformer.

To overcome the RNN limitations, Long Short-Term Memory (LSTM) [42] networks were introduced. LSTMs featured a more intricate architecture, incorporating memory cells for selective information retention and forgetting, as shown in Figure 4b. LSTMs mitigated the vanishing gradient issue and significantly enhanced long-term dependency handling in sequence models. Another notable development is the Gated Recurrent Unit (GRU) [43], designed to simplify the LSTM structure while preserving its advantages. GRUs combine the forget and input gates into a single “update gate”, and merge the cell state and hidden state, shown in Figure 4c. GRUs present a more streamlined and less complex alternative to LSTMs.

Notably, the most significant paradigm shifts in sequence modeling emerged with the introduction of the transformer model, as shown in Figure 4d. Represented in the important paper “Attention Is All You Need” [44], transformers diverged from the traditional recurrence-based approaches entirely. Instead of relying on sequence recurrence, a transformer model applies the self-attention mechanisms to process sequences. This approach assigns varying degrees of significance to different parts of the input data, enabling the model to focus on relevant portions of the sequence for prediction. Transformers demonstrate exceptional performance in various tasks, particularly excelling in handling long sequences and parallelizing computations. However, they require substantial computational resources and large datasets.

In the context of landslide studies, sequence models are mainly used for the temporal data analysis. In applications such as landslide susceptibility mapping, these models leverage multi-source data, treating it as a sequential input that encompasses satellite imagery, geological data, and other relevant information. Assimilating data from diverse sources, they can capture intricate dependencies and patterns associated with landslides. For instance, the satellite imagery allows for mapping specific topographical changes that indicate potential landslide-prone areas. Geological data aid in discerning the ground composition, enhancing the models’ ability to predict susceptibility. For the landslide displacement prediction, sequence models excel their capabilities by training on historical displacement series. By scrutinizing past displacement patterns, these models could understand temporal evolutions of ground conditions. This learning process contributes to the accuracy and reliability of predictions related to potential displacements.

#### 2.2.4. Generative Models

The evolution of generative models in machine learning exemplifies the pursuit to replicate and understand complex patterns in real-world data. This section introduces three seminal generative models: Generative Adversarial Networks (GANs) [45], Variational Autoencoders (VAEs) [46], and Diffusion Models [47], each representing a significant advancement in AI and addressing unique data generation challenges.

Introduced by Ian Goodfellow et al. in 2014, GANs emerged as a response to the need for sophisticated, high-quality data generation, particularly in image synthesis. They consist of two neural networks: the generator, which aims to produce data indistinguishable from real data, and the discriminator, tasked with differentiating between real and generated data, as shown in Figure 5a. GANs have revolutionized realistic image generation but face challenges like training instability and mode collapse, where the variety of outputs generated is limited.

Variational Autoencoders (VAEs), proposed by Kingma and Welling in 2013, address the challenge of understanding and encoding data distributions. Unlike GANs’ direct data generation approach, VAEs learn the data distribution and sample from it. They consist of an encoder, mapping input data to a latent space, and a decoder, reconstructing data from this latent representation, as shown in Figure 5b. VAEs are especially useful in applications that demand an understanding of latent data structures, such as semi-supervised learning, although they tend to produce slightly blurred outputs.

Diffusion Models, the latest innovation in generative modeling, focus on creating high-fidelity images through a unique, gradual process that transforms noise into a repre-

sensation of the target data distribution. This process includes a forward phase of adding noise and a reverse phase of removing noise to regenerate the data, as shown in Figure 5c. While Diffusion Models excel in generating detailed and diverse outputs, their high computational demands and extended training and generation times pose limitations to their scalability and practical applications in some scenarios. A comparative analysis of three types of learning models is shown in Table 2.

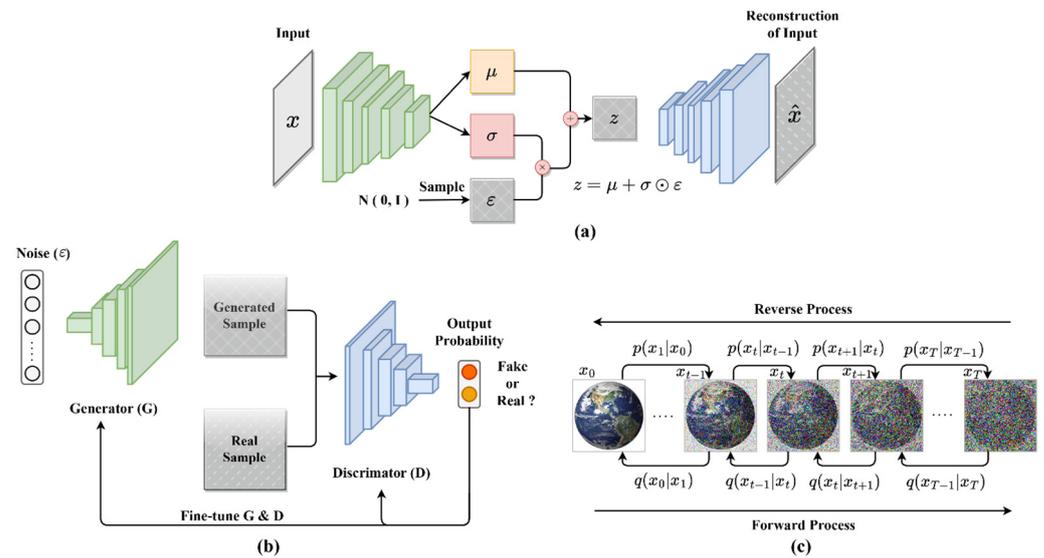


Figure 5. Architectural diagrams of generative models: (a) GAN; (b) VAE; (c) diffusion model.

Table 2. Comparative analysis of spatial, sequence, and generative models in deep learning.

Model	Classic Networks	Advantages	Limitations	Application
Vision Models	CNN architectures (LeNet, VGG, ResNet, GoogleNet)	Effective in handling spatial hierarchies, robust feature extraction	High computational resources, overfitting in deep networks	Image recognition, object detection, image segmentation
Sequence Models	RNN, LSTM, GRU, Transformer	Handles time-series data, captures long-term dependencies	Vanishing gradients in RNNs, computational intensity in LSTMs/GRUs, transformers need large datasets	Natural Language Processing, time series data analysis
Generative Models	GAN, VAE, Diffusion model	High-quality data generation, deep data representations	Training instability, complexity in model architecture	Image generation, data augmentation, anomaly detection

For landslide studies, generative models aim to addressing challenges related to data scarcity and imbalances. Specifically, these models are mainly used for data augmentation, where synthetic examples are generated to supplement existing datasets. This is particularly valuable for landslide studies where obtaining a large and diverse dataset is often challenging due to the infrequent occurrence of landslides and the diverse contributing factors. In this case, generative models not only enhance the quantity of available data but also improves the generalization capabilities of deep learning models, which is crucial for developing models that can effectively identify, classify, and predict landslides activities in regions with limited observations.

### 2.3. Loss and Optimizer

In machine learning, loss functions and optimizers are two essential components of the training process. The loss function measures a model’s performance by quantifying the error between its predictions and actual data, guiding the model towards accurate

predictions. Optimizers are algorithms that aim to minimize this loss by adjusting the model's parameters. This dynamic interplay between loss functions and optimizers is what drives the learning process of a model.

### 2.3.1. Loss

Loss functions vary based on different learning tasks. We will go over several commonly used loss functions in different scenarios. For classification tasks, which involve categorizing inputs into distinct classes, cross-entropy (CE) loss [48] is commonly used:

$$CE = -\sum_i y_i \log(p_i) \quad (6)$$

where  $y_i$  is the actual label and  $p_i$  is the predicted probability. CE loss is particularly effective in classification tasks as it penalizes incorrect classifications more significantly, driving the model towards precise classification. However, its sensitivity to dataset imbalances can be a challenge. In landslide studies, the CE loss is prevalent in landslide feature classification tasks, where its purpose is to discern features extracted from sources like satellite imagery or geological data, determining their association with landslides. It can be applied in both object-level tasks, like landslide detection, and in pixel-level tasks like precise landslide mapping. Moreover, it is also widely employed in determining the likelihood of the occurrence of a landslide event, which often refers to landslide susceptibility mapping.

In regression tasks, which predict continuous values, mean squared error (MSE) and Mean Absolute Error (MAE) are commonly used:

$$MSE = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2, MAE = \frac{1}{n} \sum_{i=1}^n |Y_i - \hat{Y}_i| \quad (7)$$

MSE highlights larger errors, while MAE provides a direct measure of error magnitude. Both are simple and interpretable, but may not fully capture complex real-world data distributions.

For segmentation tasks, particularly in medical imaging, Dice Loss [49] is always adopted for evaluating the overlap between predicted and true segmentation areas:

$$Dice = 1 - \frac{2|X \cap Y|}{|X| + |Y|} \quad (8)$$

where  $X$  and  $Y$  represent the predicted and actual segmentation areas, respectively. Dice Loss effectively addresses the problem of class imbalance and emphasizes spatial accuracy. In landslide studies, Dice Loss is often performed in the pixel-level classification task of landslide mapping. This task involves precisely delineating the boundaries of landslide-affected areas. Notably, Dice Loss is skilled of capturing fine-grained details at the pixel level, offering a distinct advantage over CE loss in spatially accurate delineation. This makes Dice Loss particularly suitable for tasks that demand intricate spatial analysis, such as landslide mapping in satellite imagery.

In image generation tasks, where the perceptual quality of the output is paramount, traditional regression losses may not be able to capture the nuanced aspects of image quality. These losses often focus on pixel-level accuracy, which does not necessarily translate to visually pleasing results. Therefore, structural similarity loss functions that better account for perceptual differences are introduced in such applications, for example, the Structural Similarity Index (SSIM) [50] and Peak Signal-to-Noise Ratio (PSNR) [51]. SSIM assesses perceptual differences between similar images, and PSNR evaluates reconstruction or compression quality. MSE is the mean-squared error between the reconstructed and original image. In landslide studies, generative loss functions, such as SSIM and PSNR, can be applied in tasks that involve the generation of synthetic remote-sensing images or maps. These loss functions are valuable when the perceptual quality of the generated outputs is highly considered. For instance, in scenarios where researchers seek to create realistic

landslide occurrence maps or simulate potential landslide events, employing such loss functions allows for a nuanced evaluation of the visual fidelity of the generated content.

### 2.3.2. Optimizer

The evolution of optimizers in machine learning has been characterized by a series of innovations, each addressing specific challenges to enhance training efficiency and performance. Stochastic Gradient Descent (SGD) [19], one of the earliest optimization algorithms, forms the foundation of many subsequent developments. It updates model parameters using the gradient of the loss function calculated from individual samples. While effective in various scenarios, its simplicity can lead to slow convergence and oscillations.

To tackle the slow convergence and oscillations characteristic of SGD, the Momentum method [19] was introduced. This method accelerates SGD by moving it along relevant directions and reducing oscillations. Momentum achieves this by incorporating a velocity component, which is influenced by past gradients, thus ensuring more consistent reduction in the loss function.

Following Momentum, AdaGrad [52] was developed to address the uniform learning rate issue in SGD. Unlike its predecessors, AdaGrad adjusts the learning rate for each parameter individually, allowing for smaller updates for frequently occurring parameters and larger updates for infrequent ones. This adaptability made AdaGrad particularly suitable for datasets with sparse features.

RMSProp [53], an improvement on AdaGrad, was designed to perform better in non-convex optimization problems, which are common in deep learning. It modifies AdaGrad's approach by using a moving average of squared gradients to normalize the gradient, thus helping the algorithm recover from steep areas and continue the learning process efficiently.

The most recent significant development is the Adam [54] optimizer, which combines the strengths of both Momentum and RMSProp. Adam maintains an exponentially decaying average of past squared gradients (like RMSProp) and an exponentially decaying average of past gradients (like Momentum). This hybrid approach allows Adam to adjust learning rates adaptively for each parameter, making it one of the most effective optimizers in various machine learning scenarios.

### 2.4. Evaluation Metrics

Model evaluation metrics are essential in the evaluation of models' prediction, providing insights into how well a model performs. Their evolution aligns with the development of machine learning, transitioning from simple metrics like error rates to more sophisticated ones. Before delving into specific metrics, it is important to introduce the confusion matrix, particularly for classification tasks. The confusion matrix is a table used to describe the performance of a classification model on a set of test data for which the true values are known (see Table 3). For binary classification, it consists of four components:

TP (True Positives): correctly identified positive cases;

TN (True Negatives): correctly identified negative cases;

FP (False Positives): incorrectly identified positive cases;

FN (False Negatives): incorrectly identified negative cases.

where, T (True) indicates correct predictions, F (False) incorrect predictions, P (Positive) the positive class, and N (Negative) the negative class.

**Table 3.** Binary classification confusion matrix.

Confusion Matrix		Actual Value	
		True	False
Predicted Value	Positive	TP	FP
	Negative	TN	FN

In classification tasks in landslide applications, such as distinguishing loess landslides from their similar surrounding environments, the fundamental objective is to assess the model's capability in accurately categorizing the data. Key metrics include Precision, Recall, Accuracy, and the F1-score [19]. Each of these metrics was developed to address specific aspects of a model's performance in classification. For instance, precision measures the proportion of true positive predictions in the positive class predictions. Recall focuses on the model's ability to identify all actual instance by measuring the proportion of true positive predictions in all actual positive instances. The F1-score, combining Precision and Recall, was introduced to provide a single metric that balances both. Accuracy calculates the proportion of correct predictions (both true positives and true negatives) out of all predictions.

$$\begin{aligned} \text{Precision} &= \frac{TP}{TP+FP}; \text{Recall} = \frac{TP}{TP+FN}; \\ \text{F1} &= 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}; \text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \end{aligned} \quad (9)$$

In addition to these metrics, for dense classification tasks such as image segmentation, geometric metrics like the Intersection over Union (IoU) are widely used. This metric quantifies the accuracy of segmentation by measuring the area overlap between the predicted and true segments.

$$\text{IoU} = \frac{\text{Area of Overlap}}{\text{Area of Union}}; \text{IoU} = \frac{TP}{TP + FP + FN} \quad (10)$$

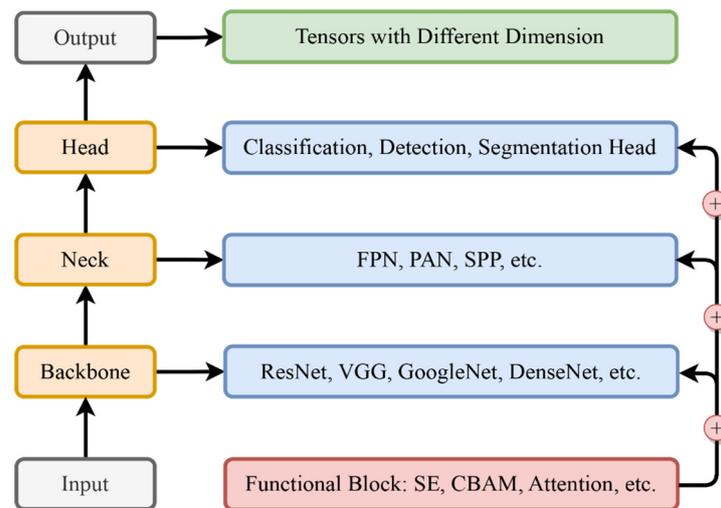
Regression tasks, which involve predicting continuous outcomes, necessitate different metrics. It is important to note that regression losses themselves can serve as metrics. These include but are not limited to metrics like the root-mean-squared error (RMSE), Mean Absolute Error (MAE), Full Reference Image Quality Assessment Metrics like the Structural Similarity Index (SSIM) [50] and Peak Signal-to-Noise Ratio (PSNR) [51], and Feature-based Image Quality Assessment Metrics like FID [55], each serving to quantify different aspects of the prediction accuracy.

## 2.5. Architecture Modules in CNN

In CNN design, the architecture is not a singular entity but a complex assembly of interconnected modules, each integral to the network's functionality and performance. At the core of this architectural framework are four primary components: the backbone, neck, head, and functional blocks. These elements, each with distinct roles and functionalities, contribute to the network's capacity for processing and interpreting complex data, as shown in Figure 6. An examination of these modules provides a comprehensive understanding of the structure of contemporary CNNs. This analysis will focus on dissecting these internal modules, elucidating their importance in forming the foundation of CNN architecture.

### 2.5.1. The Backbone: The Core of Feature Extraction

The backbone is the foundational module of a CNN architecture, primarily responsible for feature extraction. These features are essential for the subsequent layers and modules, which rely on them to make accurate predictions or classifications. The evolution of backbones reflects a journey towards deeper and more intricate networks, aiming to capture a wide range of features from the input data. Initially, the backbone module was relatively simple. Early designs like LeNet [31] were effective for less complex tasks, signifying the inception of feature extraction design. As the field developed, architectures evolved into more sophisticated structures. Networks like VGG [56], ResNet [32], and Inception [57] emerged, representing a significant leap in the backbone's design. These advanced structures were capable of capturing a broader spectrum of features from input data, laying a robust foundation for subsequent processing stages.



**Figure 6.** Architecture modules and interactions.

### 2.5.2. The Neck: Bridging and Refining Features

The neck module in CNN serves as a crucial link between the backbone and head modules, playing a vital role in refining and reorganizing the extracted features. It is instrumental in tasks requiring a detailed understanding of spatial relationships and feature hierarchies, like object detection and segmentation, e.g., landslide detection and mapping. The primary function of the neck is to enhance the network's spatial awareness by integrating and reorganizing features from various layers of the backbone, providing comprehensive and contextually relevant features to the head module. This process ensures the accuracy and reliability of the final output. Notable implementations of the neck module include the Feature Pyramid Network (FPN) [58], Path Aggregation Network (PAN) [59], and Spatial Pyramid Pooling (SPP) [60], known for their effectiveness and wide adoption.

### 2.5.3. The Head: Tailoring to Specific Tasks

The CNN's head module is intricately designed to suit specific tasks like classification, detection, or segmentation, varying based on spatial resolution and channel dimension needs. For example, in classification tasks, the head module often includes fully connected layers thanks to their effectiveness in summarizing and interpreting the features extracted by previous layers, leading to a final prediction. This is essential for classifying images into distinct categories. In detection tasks, region proposal networks (RPNs) [35] are used in the head module for their proficiency in identifying and localizing various objects within an image, crucial for tasks that require pinpointing specific items. Similarly, in segmentation tasks, the head module might integrate convolutional layers designed to perform pixel-wise classification, essential for delineating the precise boundaries of different objects in the image. Such adaptability in design ensures that the module aligns with the unique requirements of each task.

### 2.5.4. Functional Blocks: The Vanguard of Enhancement

The development of functional blocks like attention, Atrous Spatial Pyramid Pooling (ASPP) [61], Squeeze-and-Excitation (SE) [62], and the Convolutional Block Attention Module (CBAM) [63] represents the forefront of network architecture innovation. These blocks can be integrated at various stages of the network to enhance overall performance. For example, attention blocks focus on relevant features while suppressing less useful ones. SPP blocks help the network achieve spatial invariance by pooling features at various scales. SE blocks adaptively recalibrate channel-wise feature responses. CBAM blocks combine channel and spatial attention mechanisms for a thorough feature refinement. The

development of these blocks was driven by the need to address the limitations in CNNs regarding feature representation and adaptability.

Understanding architecture modules, including the backbone, neck, head, and functional blocks, is important for comprehending the design principles behind contemporary CNNs. In landslide studies, classification, detection, and segmentation are essential. The backbone, responsible for feature extraction, lays the groundwork by capturing essential information from remote-sensing imagery or geological data. Subsequently, the neck module refines and organizes these features, enhancing the spatial awareness, which is crucial for tasks like detecting and mapping the landslide-prone areas. Tailored to specific tasks, the head module aligns with the unique requirements from landslide classification, detection, or segmentation, offering adaptability to address the complexities of landslide features. The integration of functional blocks further enhances overall performance, providing a flexible approach to feature representation. As classical networks in deep learning may not always meet the specific requirements of real landslide tasks, a deep understanding of these modules empowers researchers to design networks tailored to practical landslide scenarios.

### 2.5.5. Implementing Strategies

In landslide research, effectively integrating deep learning requires a systematic approach tailored to the unique aspects of data. Initially, this involves choosing the right model architecture to suit the particular type of landslide data being analyzed. CNNs are typically preferred for analyzing spatial information in remote sensing images, thanks to their ability to handle complex spatial details. When dealing with time-series data that represent sequenced information associated with active landslides, models with recurrent layers, such as RNNs or LSTMs, are more apt given their strength in temporal data analysis. Researchers often employ robust frameworks like PyTorch, which offers extensive libraries for building and experimenting with models efficiently, making it a popular choice for developing and testing the networks.

Besides the model construction, effective management of the dataset is also important to the success of these models. It typically involves dividing the data into three segments, for example: 80% for training to establish learning patterns, 10% for validation to fine-tune the model, and the remaining 10% for testing to evaluate the network performance in an unbiased manner. Training deep learning models is an intricate process that goes beyond feeding data into algorithms. It requires thoughtful choices about loss functions and optimization algorithms, which are pivotal for guiding the learning process. Iterative hyperparameter tuning is also essential to refine and enhance model performance. Integral to avoiding overfitting are strategies like dropout and regularization; these not only prevent the model from memorizing the training data but also bolster its ability to generalize to new, unseen data.

The task of optimizing hyperparameters is particularly crucial and involves a careful balance to find the most effective model settings. Techniques such as grid search, random search, and Bayesian optimization are commonly used to explore the hyperparameter space. Regular validation is a key part of this process, ensuring that the model remains accurate and reliable. Early stopping is one such validation technique, which halts the training when improvement plateaus, preventing the potential of overfitting and helping to maintain the model's ability to generalize [19]. By adhering to these outlined steps, researchers are equipped to develop and deploy robust deep learning models using platforms such as PyTorch. This systematic approach facilitates the navigation of the complexities involved in landslide studies, thereby enhancing the accuracy and reliability of research outcomes.

## 3. Overview of Deep Learning Frameworks

In this chapter, we provide an overview of five prevalent deep learning frameworks, each tailored for distinct tasks, including three spatial tasks—classification, detection, and segmentation—a temporal task—sequence framework—and a spatial-temporal task—hybrid framework. Each framework encompasses a range of representative models de-

signed with specific architecture to meet unique requirements. Our focus is on reviewing their development, summarizing the essence of these networks within each framework, and highlighting their common structural and functional elements to offer a comprehensive understanding of their application in various contexts.

### 3.1. Deep Learning Classification Framework: The Bedrock of Feature Identification

Classification is a fundamental and essential task in computer vision. In landslide studies, classification tasks involve discerning features that are relevant to landslides, which encompasses activities like landslide susceptibility mapping. Classification networks typically encode the input into a  $1 \times 1 \times N$  tensor, condensing spatial dimensions to a single point. This representation enables the network to make high-level predictions, such as identifying landslide presence or assessing the region susceptibility. Over years, these architectures have significantly evolved, influencing advancements in related vision tasks such as detection and segmentation, primarily through the shared utilization of backbone for effective feature extraction. The architectural diagram of the classification framework is shown in Figure 7.

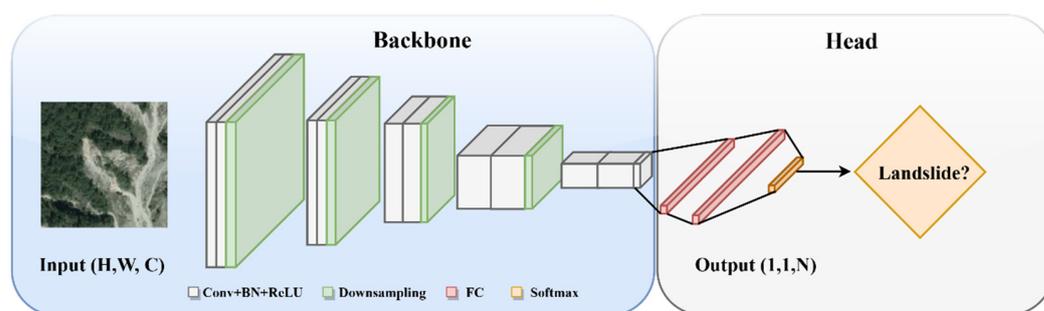


Figure 7. Architectural diagram of classification framework.

The evolution of classification networks in deep learning began with LeNet [31], a groundbreaking CNN model that set the stage for neural network applications in visual data processing. This evolution was propelled further by pivotal models like AlexNet, which showcased the efficacy of CNNs in image classification tasks. Subsequently, VGG [56] took a leap forward by exploring very deep architectures with small  $3 \times 3$  convolutional filters, which further enhanced the network's performance. ResNet [32], another milestone, innovatively addressed the vanishing gradient problem through its residual connections. These foundational networks not only solidified the core principles of CNNs but also markedly improved the capability to discern and categorize complex image patterns.

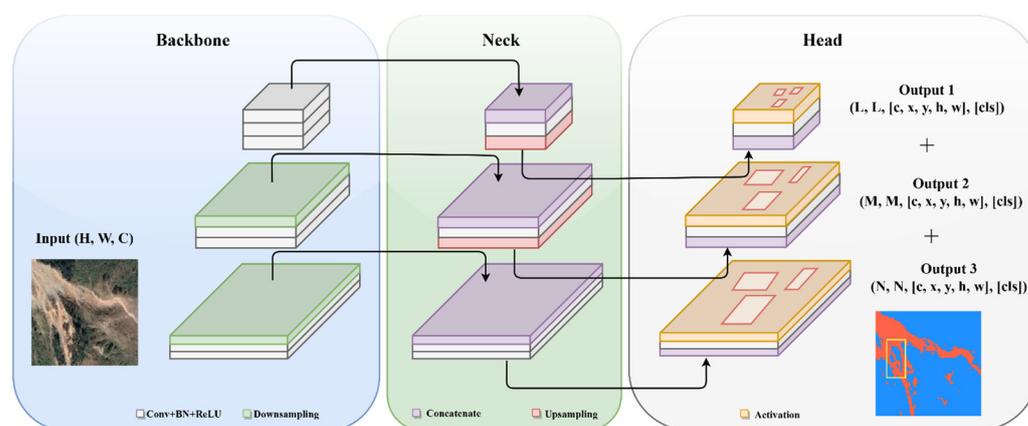
Simultaneously, exploration emerged to broaden network architectures. Models like ResNext [64] utilized grouped convolutions to expand the network's structure, enhancing its capacity without a substantial increase in computational demands. In parallel, GoogleNet's Inception modules [57] made strides in multi-scale feature extraction, showing improved results over previous architectures. Concurrently, lightweight networks like EfficientNet [65] and ShuffleNet [66] were proposed for working in resource-constrained settings, such as on mobile devices. More recently, the Vision Transformer (ViT) [67] gained huge attention, marking a transition in backbone architectures. Diverging from its predecessors, ViT adopts the transformer architecture, commonly used in Natural Language Processing (NLP), for image processing. This model has demonstrated notable results, particularly in data-rich environments, signifying a trend from traditional convolution-based methods to attention-based feature extraction.

For classification networks, the main focus is on devising the backbone architecture for effective feature extraction. This foundational work has facilitated their application in other spatial tasks, such as detection and segmentation. Generally, the neck component in these networks is integrated with the overall structure, smoothly transitioning from feature extraction to classification. The head of these networks consists of fully connected layers,

often ending in a SoftMax layer (Figure 7), which is responsible for projecting the extracted features into a probability distribution across different classes. The output of these networks corresponds to the number of predefined classes. In such frameworks, cross-entropy is a commonly used loss function due to its efficacy in assessing classification accuracy.

### 3.2. Deep Learning Detection Framework: Balancing Localization and Identification

Object detection is a crucial task in computer vision, distinct from simple image classification. In landslide studies, object detection is naturally involved in detecting and locating specific objects or features relevant to landslides. This encompasses the detection of potential landslides, precursor deformation, and other relevant elements within remote-sensing images or geological data. Object detection involves identifying and localizing multiple objects within an image, which requires a sophisticated design for its neck and head modules to handle the complexity. Consequently, a variety of architectural styles, such as two-stage, one-stage, anchor-free, and transformer-based detectors, have been developed. Each of these architectures has its unique approach, with varying strengths and methods to fulfill the requirements of object detection. The architectural diagram of the detection framework is shown in Figure 8.



**Figure 8.** Architectural diagram of detection framework.

The development of object detection frameworks has undergone significant advancements, starting with the two-stage RCNN series [33–35], including RCNN, Fast RCNN, and Faster RCNN. These frameworks focus on generating region proposals for potential objects in an image and then classifying each region. While highly accurate, they are computationally intensive due to the two-step process. Seeking to balance speed and accuracy, one-stage detectors like the “Single Shot MultiBox Detector (SSD)” [68] and the “You Only Look Once (YOLO)” series [36–38] emerged. These models streamline the detection process, directly predicting object classes and locations in a single forward propagation of the network. While offering faster detection speeds, they sometimes compromise on accuracy compared to two-stage models, particularly in complex scenes or with small objects.

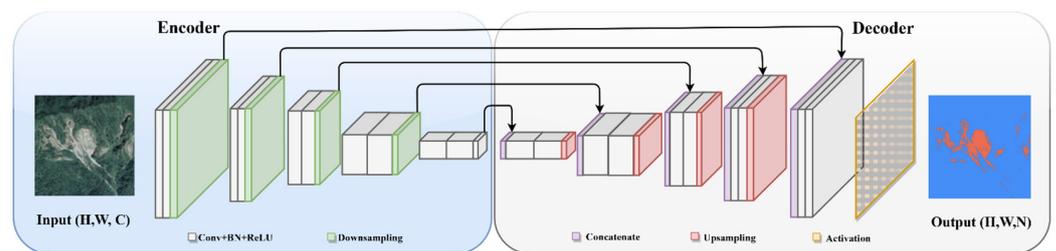
In parallel, efforts to transcend the constraints of traditional anchor-based approaches have led to the rise of anchor-free methods like CenterNet [69] and CornerNet [70]. These innovative techniques localize objects without depending on predefined anchor boxes, thus simplifying the model and reducing computational load. Alongside this, the involvement of transformer-based techniques, notably exemplified by DETR (Detection Transformer) [71], has significantly altered the landscape of object detection. The DETR leverages the power of transformers to improve object detection by effectively handling complex spatial relationships and long-range dependencies. This shift towards transformer-based methods represents a major evolution in object detection, offering a new perspective on handling spatial data complexities.

In the evolution of detection frameworks, two crucial components are the neck and head. The neck, often incorporating structures like Feature Pyramid Networks (FPN) [58]

and Path Aggregation Network (PAN) [59], fuses multi-scale features from various layers of the backbone. This fusion is pivotal for detecting objects at different scales, merging vital information for both object localization and classification. The head has a dual function: localizing objects, typically via bounding box coordinates, and classifying them. To meet these dual requirements, a composite loss function, usually a combination of localization loss (such as Smooth L1) and classification loss (cross-entropy), is employed. This blended loss is fundamental for object detection frameworks, ensuring precise object classification and accurate spatial localization.

### 3.3. Deep Learning Segmentation Framework: The Intricacy of Pixel-Level Classification

The segmentation framework, characterized by pixel-wise classification, is inherently more complex than the classification framework due to its need for detailed spatial analysis at the pixel level. In landslide studies, segmentation is applied to delineate the precise boundaries of landslide-affected areas, aiding in a detailed and accurate spatial analysis. It is particularly useful for tasks such as landslide mapping, where the goal is to classify each pixel within an image to accurately identify and map areas impacted by landslides. The backbone in segmentation framework is also aimed at feature extraction, and commonly has similar choices like ResNet and VGG that are used in classification networks. Of course, these backbones might undergo slight modifications such as layer adjustments or block addition, to fit the specific requirements of the practical task. Unlike detection frameworks, the role of the neck in segmentation is not always distinct and it is often integrated within the backbone. Notable examples include the Atrous Spatial Pyramid Pooling (ASPP) [61], Pyramid Scene Parsing Network (PSPNet) [72] and Spatial Pyramid Pooling (SPP) [60], which are designed to enhance multi-scale feature awareness, enabling an accurate dense prediction. The architectural diagram of segmentation framework is shown in Figure 9.



**Figure 9.** Architectural diagram of segmentation framework.

Distinguished from classification framework, segmentation networks aim for dense predictions, mostly necessitating output dimensions matching the input resolution, often represented as  $H \times W \times N$ , where  $H$  and  $W$  denote height and width, respectively. This structure is ideal for pixel-level tasks like landslide mapping, as it provides detailed spatial information. Therefore, after down-sampling via the backbone, it becomes crucial to up-sample the features back to their original resolution, requiring the use of a decoder. The evolution of decoders parallels the development of classic segmentation networks. This began with the Fully Convolutional Network (FCN) [39], which introduced transposed convolution layers for up-sampling, enabling end-to-end training and precise pixel-level predictions. Enhancements followed with FCN variants like FCN-8s, which integrated skip connections. These connections fused deep, coarse semantic information with shallow, fine appearance details, thereby augmenting spatial precision, particularly around object boundaries.

This progression continued with the advent of U-Net [40], which introduced a symmetric encoder–decoder architecture, where the decoder progressively up-samples feature maps while concatenating them with corresponding encoder feature maps, significantly improving localization capability. Following this, the DeepLab series (v1–v3+) [73–75] brought novel advancements. They introduced atrous (dilated) convolutions in both the encoder and decoder stages and incorporated Atrous Spatial Pyramid Pooling (ASPP) in

the encoder, alongside a new decoder design focused on refining segmentation outcomes and edge clarity. Later developments, such as the Attention U-Net [76], expanded these innovations by embedding attention mechanisms within the decoder. This integration directed the model's focus toward more relevant regions, and further enhanced precision in segmentations across diverse scenarios. Most lately, transformer-based models [77] have applied self-attention to capture long-range dependencies across the entire image, providing a global context for segmentation. These advancements indicate a trend towards networks that not only understand local features but also incorporate a wider image context, leading to more accurate and detailed segmentation results.

In segmentation frameworks, the choice of loss function can vary depending on the specific requirements of the task. Commonly used loss functions include pixel-wise cross-entropy, which is effective for classifying each pixel individually. Additionally, geometric loss functions like Dice Loss or IoU loss are employed to accurately capture the shapes and overlaps of segmented areas [49]. These loss functions can be used individually or in combination, depending on the need to balance pixel-level accuracy with the geometric representation of segmentation results.

### 3.4. Deep Learning Sequence Framework: Contextual Data Modeling

Sequence data analyzing in deep learning represents a unique and challenging domain, distinct from the realms of classification, detection, and segmentation. It focuses on the temporal dynamics and dependencies within the contextual data, a critical aspect in a wide range of applications. In landslide studies, sequence data comprise a diverse input from multi-sources, including remote sensing, geo-environmental, meteorological, and topographical data, etc. Additionally, it also involves time-series data, such as remote-sensing images, and historical pointwise deformation measurements. This section delves into the evolution of deep learning sequence frameworks, tracing their development from initial models to the sophisticated networks we see today. The architectural diagram of sequence framework is shown in Figure 10a.

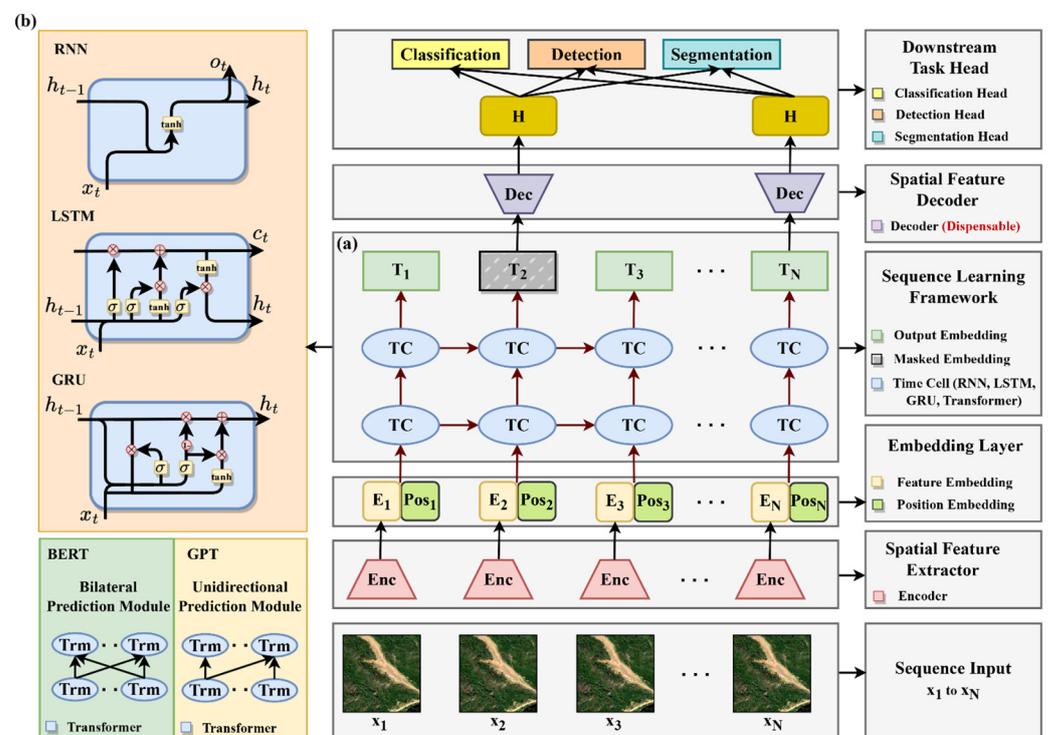


Figure 10. Architectural diagram of hybrid framework: (a) sequence framework; (b) hybrid framework.

Being the initial representative work, Recurrent Neural Networks (RNNs) [41] provide a fundamental approach to sequence processing by integrating memory elements into

neural networks. However, RNNs have encountered limitations in managing long-term dependencies, which led to the development of more advanced architectures. Long Short-Term Memory (LSTM) [42] networks overcame these challenges by incorporating gates that control information flow, thereby adeptly capturing extended dependencies in sequences. Subsequently, Gated Recurrent Units (GRUs) [43] were introduced as a more efficient variant, delivering similar functionalities to LSTMs but with a less complex architecture, which reduced computational requirements. A pivotal shift occurred with the involvement of the transformer model [44], which moved away from reliance on recurrent mechanisms. Leveraging self-attention mechanisms, transformers are capable of processing entire sequences in parallel, thus effectively capturing dependencies between sequence elements that are widely separated by many positions, surpassing the capabilities of their predecessors. This architecture has since become a foundational element, inspiring numerous subsequent innovations in sequence modeling.

The recent advances in deep learning for sequence data, especially in Natural Language Processing (NLP), have been characterized by employing transformer architecture and extensive data training. This transition began with the development of ELMo (Embeddings from Language Models) [78]. Although ELMo does not use transformer architecture, it laid the groundwork for creating deep, contextualized word representations. ELMo's bidirectional LSTM structure was instrumental in generating word embeddings that comprehensively capture semantic attributes from surrounding text, paving the way for the later adoption of transformer-based models in NLP. This evolution was further propelled by the development of BERT (Bidirectional Encoder Representations from Transformers) [79] by Google. BERT, utilizing the transformer architecture, introduced an innovative bidirectional training method. This approach allows the model to contextualize information from both directions of a sentence, leading to a more nuanced understanding of language. BERT's effectiveness in tasks such as question answering and language inference is largely due to this bidirectional processing, coupled with its training on extensive text datasets. Simultaneously, OpenAI's GPT (Generative Pretrained Transformer) series [80–83], particularly GPT-4, showcased the extraordinary effectiveness of transformer-based models on a large-scale language dataset. Focusing primarily on unsupervised learning and trained on a diverse array of internet text, GPT models have demonstrated a remarkable ability to generate coherent and contextually relevant text sequences. The success of the GPT series has highlighted the capability of large transformer-based models to grasp and reproduce the intricacies of human language, marking a significant milestone in the field of NLP.

### 3.5. Deep Learning Hybrid Framework and Transfer Learning

As reviewed in the previous sections, the spatial frameworks are skilled at extracting and interpreting complex spatial patterns present in spatial data, adeptly handling tasks such as classification, detection, and segmentation. In contrast, sequence frameworks excel at processing and comprehending time-series data. Thus, combining these two frameworks to foster a hybrid framework, we can perform more complex and advanced monitoring and prediction tasks, which have the potential to enhance the accuracy and reliability of predictive results, and to improve the overall understanding in landslide dynamics. The architectural diagram of hybrid framework is shown in Figure 10b.

In such an integrated hybrid framework, the backbone of the spatial network primarily extracts spatial features from observational data collected at each time point. These features are then transformed into embeddings, configured to meet the sequence framework's input requirements. This process ensures that the spatial characteristics of the data are captured comprehensively. Subsequently, these spatial embeddings are processed through a sequence framework to learn the contextual relationships. This step is crucial for uncovering latent connections and temporal patterns within the data. A variety of architectures, ranging from traditional models like RNN and LSTM to advanced models such as BERT and GPT, can be employed for this contextual learning. The choice of architecture depends on the complexity of the task and the nature of the data. Once the temporal relationships are

established, the processed features are then fed into the corresponding head of the spatial framework. Here, the head can be selected corresponding to specific tasks at hand. For instance, we can install the classical detection heads, like the YOLO series, to conduct a landslide detection task, while for a landslide displacement prediction task, we typically adopt classification heads with regression loss to implement. This modular approach, separating the spatial framework into functional components like the backbone, neck, and head, and integrating them with sequence frameworks, provides high flexibility and convenience. It allows for the seamless assembly and reconfiguration of various network structures to meet the specific needs of diverse real-world tasks.

Notably, training such a model effectively requires a strategic approach, and transfer learning presents a viable solution. Initially, the spatial network can be trained using data acquired at a single time point. Once this training is complete, the parameters of the backbone and head will be frozen, and then they can be integrated with the chosen sequence framework to train the sequence framework individually. Finally, we can fine-tune the entire integrated system—both spatial and temporal components—to ensure the optimal training performance. This innovative integration of spatial and temporal frameworks offers an advanced tool for both real-time monitoring and predictive modeling. A comparative analysis of five prevalent deep learning frameworks is shown in Table 4.

**Table 4.** Comparative analysis of deep learning frameworks.

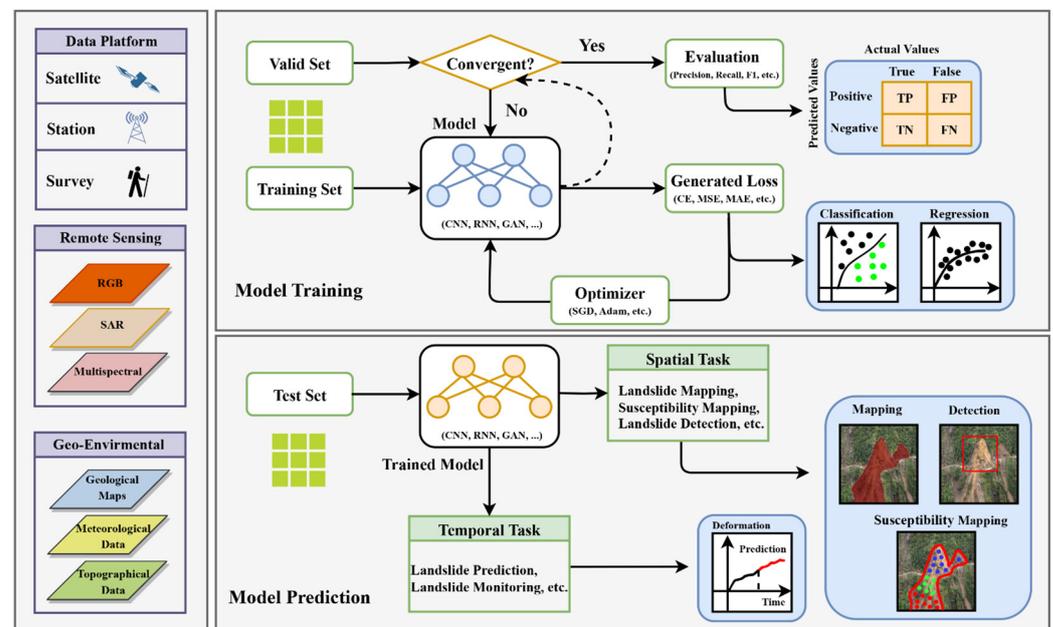
Framework	Backbone	Neck	Head	Output	Loss
Classification Framework	CNN (e.g., ResNet, VGG)	Integrated	Fully connected layer	Class numbers	Cross-entropy
Detection Framework	CNN (e.g., ResNet, VGG)	FPN, PAN	RPN (two-stage), Direct Predict (one-stage)	Box coordinates, Class probs	Localization (e.g., L1/L2) + Classification (Cross-entropy)
Segmentation Framework	CNN (e.g., ResNet, VGG)	ASPP, Dilated convolutions	Convolutional layer	Per-pixel class labels	Geometric (e.g., Dice, IoU), Cross-entropy
Sequence Framework	RNN (e.g., LSTM, GRU), Transformer	Not distinct	Not distinct	Variable (Time series, Sequences)	MSE, Cross-entropy
Hybrid Framework	CNN + RNN/Transformer	Application-specific	Application-specific	Application-specific	Combined spatial-temporal losses

## 4. Deep Learning Frameworks Application for Landslides

### 4.1. Landslides

Landslides are among the most common and destructive natural disasters in mountainous areas, posing significant threats to human lives and properties. Alongside, they can cause damages to critical infrastructures like roads, bridges, and power lines [84]. They also lead to the destruction of vegetation, soil erosion, and land degradation. The frequency and impact of landslides depend on complex environmental and triggering factors [85]. Environmental factors such as soil composition and geological features play a crucial role in determining the possible locations of landslides. On the other hand, triggering factors like climatic changes and seismic events are pivotal in initiating these events. The intricate and nonlinear interplay of these factors challenges traditional methods for modeling their physical mechanisms and predicting their development, which has prompted researchers to pivot towards more effective, data-driven approaches [86]. Recent advancements in deep learning have shown significant capabilities in extracting landslide-related information from large datasets [87]. The amalgamation of deep learning with extensive remote sensing from satellites and in situ sensors has equipped researchers with a powerful tool for landslide research and prevention [88]. Efforts mainly can be divided into two aspects.

The first aspect focuses on known landslides, in which researchers employ multi-source data, including remote sensing data (SAR, optical, and multispectral sensors) and geo-environmental data (such as geological, hydrological, meteorological, and topographical data.) to predict potential landslides and assess their risk levels. This process is known as susceptibility mapping [89]. Another key effort is displacement prediction, which uses historical data on landslide movement and related environmental factors to forecast their future displacements [90]. Regarding the second aspect, scholars have aimed at addressing slow-moving landslides, and have developed algorithms for landslide detection and mapping from remote sensing images, each serving distinct yet complementary purposes. Landslide detection, taking an object perspective, uses algorithms to locate landslides and approximate their areas with bounding boxes [91]. This method is particularly effective for quickly identifying landslide locations in extensive surveys. In contrast, landslide mapping, adopting a pixel-based approach through segmentation techniques, offers a more granular analysis. It precisely defines the boundaries of each landslide at the pixel level, providing vital details for comprehensive risk assessments and mitigation strategies [92]. General data processing and the workflow using deep learning for landslide detection, mapping, susceptibility mapping, and displacement prediction is shown in Figure 11.



**Figure 11.** Deep learning workflow for landslide studies.

In the following section, we will investigate the various deep learning frameworks suitable for each of four major tasks in landslide studies: landslide detection, mapping, susceptibility mapping, and displacement prediction. For each task, we highlight specific deep learning frameworks that are applicable and review related studies that have utilized these frameworks. We aim to introduce a clear perspective on the selection of frameworks for each task and provide insights into the latest achievements and performances of these techniques in the context of landslide risk management.

#### 4.2. Landslide Detection (Object-Based)

Deep learning has revolutionized the field of remote sensing, particularly in challenging tasks such as landslide detection. This section delves into landslide detection using object-based deep learning approaches, e.g., a detection framework. Generally, a detection framework includes two-stage, one-stage, anchor free, and transformer-based detectors. A thorough review of the literature shows that most of its application in landslide detection is in the form of primary two-stage [33–35] and one-stage detectors [36–38],

while anchor-free and transformer-based detectors are less common. This could be due to anchor-free detectors [69,70] and transformers [71] being new and not as widely applied. Here, we will explore the latest developments and applications in these technologies for landslide detection.

Two-stage detection networks typically involve a region proposal network (RPN) for generating candidate bounding boxes and then refine and classify these regions in a second stage, leading to potentially higher accuracy but slower inference times. Representative networks include RCNN, Fast RCNN, and Faster R-CNN [33–35]. Applied them into landslide detection, Guan et al. (2023) [93] used an improved Faster R-CNN to detect slope failures and landslides in the Huangdao District of Qingdao in China, where they introduced a multi-scale feature enhancement module into Faster R-CNN to enhance the network's perception of different scales of landslides. The experiments showed that the improved model outperformed the traditional version, with an AP of 90.68%, F1-score of 0.94, recall of 90.68%, and precision of 98.17%. The authors applied it to detect geological hazards of slope failure in Huangdao District, and it only missed two landslides, demonstrating a high detection accuracy. To improve the accuracy of landslide detection in satellite images, Tanatipuknon et al. (2021) [94] combined two object detection models based on Faster R-CNN with a classification decision tree (DT). In detail, the first Faster R-CNN was trained on true color (RGB) images, and the other was trained with grayscale DEMs. The results from both models were employed by a DT to generate bounding boxes around landslide areas. Compared to those using either RGB or DEMs, this integrated approach showed improved performance in various evaluation metrics, demonstrating its effectiveness in landslide detection.

Due to the large difference in the landslide scale and the important characteristics of the landslide's shallow layer, Zhang et al. (2022) [95] modified the original Faster R-CNN by incorporating the Feature Pyramid Network (FPN) with ResNet50. This integration aimed to address the challenges posed by the varying scales and shallow layer characteristics of landslides. Validated with a public landslide inventory from Bijie City, Guizhou, China, the modified model demonstrated a significant improvement in accuracy (87.10%) compared to that of VGG16 (70.20%), which indicates a promising method for future research. The InSAR technique has immense potential for detecting active landslides by its unique advantage in measuring subtle ground deformation. However, the operational application of InSAR for landslide detection in a wide area is still hindered by the high labor and time costs for the visual interpretation and manual editing of the InSAR-derived velocity maps. Aiming at this obstacle, Cai et al. (2023) [96] developed a method combining InSAR and a CNN, specifically using an improved Faster RCNN with attended ResNet-34 and FPN. Applied in Guizhou province, China, this approach successfully identified over 1600 active landslides from a substantial number of Sentinel-1 and PALSAR-2 images, demonstrating high precision and recall in various test areas. This method shows significant promise in efficiently updating landslide inventories and aiding in disaster prevention. To quickly detect landslide hazards for a timely emergency rescue, Yang et al. (2022) [97] proposed an enhanced Faster R-CNN for landslide detection. Specifically, this method has several aims, including image quality improvement, batch size elimination using group normalization, multiscale feature fusion with a Feature Pyramid Network, and employing a deep residual shrinkage network as the backbone to extract complex spatial features. Experimental results indicate a notable improvement in accuracy and average precision compared to the standard Faster R-CNN and other one-stage models, such as YOLOv4 and SSD, which proves the model's effectiveness in landslide detection.

Different from two-stage detectors, one-stage detection networks directly predict bounding boxes and class scores in a single step using predefined anchor boxes, which makes them faster, simpler, and suitable for real-time applications. Representative networks include the YOLO series [36–38], SSD [68], and RetinaNet [98]. For their application in landslide detection, Fu et al. (2022) [99] presented a novel method for detecting slow-moving landslides from stacked phase-gradient maps, aiming at overcoming the limitations

of phase unwrapping errors and atmospheric effects. In detail, they developed a burst-based, phase-gradient stacking algorithm to sum up phase gradients in short-temporal-baseline interferograms, then trained an Attention-YOLOv3 network with manually labeled landslides on the stacked phase gradient maps to achieve a quick and automatic detection. Applying this method to an area of approximately 180,000 km<sup>2</sup> in southwestern China, they identified 3366 slow-moving landslides. By comparing these results with optical imagery and previously published landslide data for the region, their method proved to be precise and efficient in automatic detection across large areas. Notably, it unveiled about 10 additional counties with high landslide density, beyond the known high-risk areas, emphasizing the need for increased geohazard attention in these locations. This finding demonstrates the potential of their method for nation-wide slow-moving landslide detection, offering a crucial tool for improved geohazard monitoring and risk management. Regarding the need of an open, large, and widely recognized landslide dataset, Wang et al. (2021) [100] created a landslide dataset through open satellite image data, in which the landslide boundary was marked by professional engineering geologists. Building upon the foundational YOLOv5, they innovatively integrated Adaptively Spatial Feature Fusion (ASFF) and the Convolutional Block Attention Module (CBAM) to enrich the model's capability to assimilate multi-scale feature information. This enhancement yielded a 1.64% improvement in model performance. Traditional field survey, while reliable for small-scale investigations, proves inadequate for large areas due to its cost and labor intensity. To cope with this limitation, an innovative approach [101] was proposed to detect and monitor these hazards in China's high-mountain areas. The researchers firstly tested the feasibility of using SBAS-InSAR with Sentinel-1A data for landslide detection in the Yunnan–Myanmar border region. Subsequently, they employed YOLOv3 with Gaofen-2 images for further analysis. Applied in Fugong County, Yunnan Province, results showed that most landslides identified by manual interpretation were detected by SBAS-InSAR as true positives, accounting for 68.75% of the total references. This indicates that the majority of these were active during the study period and posed potential threats to the surrounding areas, underscoring the importance of such advanced detection methods for public and local authority awareness.

Two-stage detectors, though accurate, tend to be slower, whereas one-stage detectors offer faster detection but at a cost to accuracy. To concur this, an improved one-stage detection model [102], YOLO-SA, for use in emergency rescue and evaluation decision-making with high-spatial-resolution remote sensing images on mobile and embedded equipment, was proposed. In this study, two enhancements were presented, including adopting group convolution, ghost bottleneck modules to reduce parameters, and integrating an attention mechanism for improving accuracy. Tested in Qiaojia and Ludian counties, Yunnan, China, YOLO-SA outperformed various advanced models in terms of parameter efficiency, accuracy, and speed, which confirms its effectiveness in potential landslide detection in near-real time. Another challenge poses itself that in some scenarios, remote-sensing satellites may not be able to timeously obtain the image from the disaster areas due to the orbital cycle and weather impacts. Addressing this limitation, Yang et al. (2022) [103] utilized UAV images from the Nepal earthquake-affected area of Zhangmu Port and the transfer learning strategy to use their detection model, which aims to overcome the shortage of sufficient training data. Comparative analysis revealed that this approach surpassed the detection performance of the SSD model, offering an effective solution for quick and accurate landslide detection.

No matter whether a two-stage or one-stage detector, each actually possess unique strengths and limitations in the context of landslide detection. Researchers have established a comparative analysis of multiple classical detectors, providing valuable insights into their efficacy in the same task. For example, Wu et al. (2022) [104] tested representative algorithms like Faster R-CNN, YOLOv3, and SSD on a substantial landslide remote sensing dataset. Technical advantages and functional characteristics of each algorithm were thoroughly compared and analyzed. Their results demonstrated that while the Faster

R-CNN algorithm offers higher accuracy, YOLOv3 and SSD are more suited for timely monitoring and practical applications due to their faster detection speeds. Similarly, Zhang et al. (2022) [105] constructed a comprehensive dataset using Google Earth imagery. They applied and evaluated the performance of various detectors including YOLOV5, Faster RCNN, EfficientNet, SSD, and a modified YOLOV5 embedded with CBAM and Ghost modules. The findings indicated that SSD was highly effective in detecting scenarios where only one landslide existed within an image. On the other hand, the modified YOLOV5 demonstrated proficiency in identifying scenarios with multiple landslide events in an image, which strikes a balance between detection capabilities and model complexity. Detecting loess landslides has always been challenging due to their similarity to the surrounding environment in optical images. To tackle this, Ju et al. (2022) [106] employed different deep learning methods for detecting loess landslides from Google Earth images. They established a database with 6111 interpreted landslides from three areas in Gansu Province, China. They evaluated three detection networks, including RetinaNet, YOLO v3, and Mask R-CNN. Their results showed that the detection accuracy is positively correlated to landslide areas; namely, larger landslides can be identified more accurately. Among the tested models, the Mask R-CNN demonstrated the highest accuracy, with an F1-score of 55.31%, which proves the viability of object detection methods in automated loess landslide detection from satellite imagery. A summary of case studies of landslide detection is shown in Table 5.

**Table 5.** Summary of case studies of landslide detection.

Study Area	Data Used	Model Evaluated	Metric and Result (%)	Reference
Huangdao District, QingDao, China	Orthographic Remote Sensing Image	Improved Faster RCNN	AP (90.68), F1-score <sup>opt</sup> (94), Recall <sup>opt</sup> (90.68), Precision <sup>opt</sup> (98.17)	Guan et al. (2023) [93]
Bijie City, Guizhou, China	TripleSat Image, Grayscale DEM	Faster R-CNN + DT Tree	AP (86), F1-score <sup>avg</sup> (79), Recall <sup>avg</sup> (77), Accuracy <sup>avg</sup> (97)	Tanatipuknon et al. (2021) [94]
Guangzhou, China	Sentinel-1 and PALSAR-2 image	Faster RCNN + ResNet-34 + FPN	AP (63.90), F1-score <sup>opt</sup> (91.40), Recall <sup>opt</sup> (91.49), Precision <sup>opt</sup> (91.33)	Cai et al. (2023) [96]
Southwestern China	Sentinel-1	YOLO-Tiny, YOLO-v3, Improved YOLO-v3 (best)	AP (75), F1-score <sup>opt</sup> (90.82), Recall <sup>opt</sup> (87), Precision <sup>opt</sup> (95)	Fu et al. (2022) [99]
Sichuan, China	Optical Image	YOLO-v5 + ASFF + CBAM	AP (74.01), F1-score <sup>opt</sup> (77.30), Recall <sup>opt</sup> (76.21), Precision <sup>opt</sup> (78.42)	Wang et al. (2021) [100]
Fugong County, Yunnan, China	Sentinel-1 and Gaofen-2 image	SBAS-InSAR + YOLOv3	MAP 50 (99.17), MAP 50:95 (73.50)	Guo et al. (2022) [101]
Qiaojia and Ludian counties, Yunnan, China	Optical Image	Faster RCNN, SSD, YOLO-v4, YOLO-SA (best)	AP (94.08) with FPS (42 f/s)	Cheng et al. (2021) [102]
Zhangmu Port, Tibet, China	UAV Image	SSD, SSD + Transfer Learning (best)	AP (95.10), F1-score <sup>avg</sup> (84), Recall <sup>avg</sup> (90)	Yang et al. (2022) [103]

<sup>opt</sup>: the optimal result during the experiment; <sup>avg</sup>: the average result across the confidence threshold from 0 to 1.

#### 4.3. Landslide Mapping (Pixel-Level)

Landslide mapping is a critical task for disaster management and mitigation. The complexity and variability of landslides make this task challenging for traditional methods. By treating landslide mapping as a pixel-level segmentation task, deep learning, with its advanced segmentation frameworks, offers a powerful solution. These frameworks can precisely and quickly delineate landslide-prone areas, improving the capability of emergent

responses [107]. Here, we explore the latest achievement of landslide mapping using deep learning segmentation frameworks, highlighting how they transform our ability to identify and manage these natural hazards.

Being a successful and classical segmentation network, UNet [40] and its variants have been widely applied in landslide mapping. For example, Bragagnolo et al. (2021) [108] employed UNet to identify landslide scars using Landsat-8 satellite data in a region of Nepal. The experimental results showed the superior accuracy of UNet over compared methods in terms of recall (74%), precision (61%) and the F1-Score (67%). This indicates the potential of UNet in dynamic mapping systems for landslide scar mapping. To accurately extract landslides from high-resolution remote sensing images, a modified version of UNet, named L-UNet, [109] was developed. This model enhanced the original UNet by incorporating a multi-scale feature-fusion (MFF) module, a residual attention block, and a data-dependent upsampling (DUpsampling) block. When applied to the northern mountainous area of Huizhou City, Anhui, China, the L-UNet demonstrated superior performance over the baseline UNet, with an increased precision (4.15%), recall (2.65%), mIoU (4.82%), and F1-Score (3.37%), proving its capability in accurately and efficiently extracting landslides from remote sensing images. InSAR is an effective tool for active landslide recognition. However, its interpretation highly relies on expert experience, which is time-consuming and subjective. To bridge this gap, Chen et al. (2022) [110] developed a Deep Residual Shrinkage UNet (DRs-UNet) for mapping landslides from SAR interferograms, or InSAR-derived velocity maps. The proposed network, enhancing the standard U-Net with a Residual Shrinkage Building Unit (RSBU), significantly reduced the noise level of InSAR images and improved feature extraction. Compared to established models like UNet and SegNet, DRs-UNet demonstrated superior performance, achieving a higher F1-Score and an impressive IoU over 90%. Tested at a landslide-prone area in Zhongxinrong County, the quantitative evaluation results showed that the DRs-UNet can effectively recognize potential landslide hazards from InSAR imagery.

In addition to classical UNet architectures, researchers also explored a state-of-the-art transformer-based network [44] for landslide mapping. For instance, Huang et al. (2023) [111] indicated that current deep learning approaches face a challenge of a long inferring time, which is inefficient for disaster prevention and post-disaster rescue. To show this limitation, they developed a novel Distilled Swin-Transformer (DST) for landslide recognition. When tested in Zigui County, Hubei, China, the DST model achieved the highest performance in various evaluation metrics compared to ResNet50, Swin-Transformer, and the DeiT model. Meanwhile, DST also demonstrated an optimal computational cost, with the lowest FLOPs, making it highly suitable for rapid landslide recognition in post-disaster efforts. Can we integrate transformers into a CNN to improve the model's capability for landslide mapping? Yang et al. (2022) [112] explored this possibility by integrating a transformer with the ResUNet, in which a spatial and channel attention module is added to the decoder for reducing the noise in feature maps. Training with small datasets, this model was expected to capture the global context feature and recognize these landslides. Applying this integrated method to Bijie region and Iburi earthquake-induced landslides, the model not only outperformed the standard ResUNet, achieving a higher mIoU and F1-score, but also detected a significant number of landslide incidents, which is useful for improving the accuracy and reliability of landslide emergency rescue operations. Similarly, Wang et al. (2022) [113] introduced a Separable Channel Attention Network (SCANet) for landslide mapping. The SCANet employs a Poolformer encoder and a Separable Channel Attention Feature Pyramid Network (SCA-FPN) decoder. The Poolformer, leveraging transformer architecture, captures global semantic information, and the SCA-FPN enhances multi-scale semantic fusion and precise pixel-level prediction. The experiment showed that the SCANet surpassed the traditional network ResNet50-Unet, achieving a higher mIoU (1.95%), although with fewer parameters.

To attain optimal results in landslide mapping, conducting comparative experiments across various segmentation models is a promising approach. Researchers have explored

the efficacy of different deep learning models to determine which offers the best performance for their specific task. A notable example [87] is the study by Ghorbanzadeh et al. (2022), where they created the Landslide4Sense benchmark dataset, a comprehensive collection of 3799 image patches combining data from Sentinel-2 and ALOS PALSAR. The team assessed eleven advanced segmentation models, including U-Net, ResUNet, and others, on this dataset. This research underscores the importance of selecting the right model and fine-tuning parameters for optimal application in real-world scenarios. Similarly, to address the pressing need for quick and effective landslide mapping in disaster rescues, Yang et al. (2022) [114] compared three detection models, including U-Net, DeepLabv3+, and PSPNet, each evaluated with different backbone networks for determining the most effective model for landslide recognition. The experiment revealed that the PSPNet with the ResNet50 backbone achieved the highest accuracy in terms of the mIoU (91.18%). Using optical remote sensing images, Du et al. (2021) [115] evaluated six segmentation models, FCN, UNet, PSPNet, Global Convolutional Network (GCN), DeepLabv3, and DeepLabv3+, on a new open-source landslide dataset. Their results showed that GCN and DeepLabv3 are most suitable for this problem scenario, with the highest mIoU and pixel accuracy. These studies highlight that the effectiveness of a model can vary across different datasets and scenarios, which emphasizes the importance of conducting experiments to select the appropriate model for practical tasks.

In addition to employing these established segmentation networks, scholars also devised their own networks for meeting the specific requirement of the task at hand. For example, mapping ancient landslides from high-resolution remote sensing images (HRSIs) poses significant challenges due to their transformed morphology over time and resemblance to surroundings. Addressing these challenges, a novel Iterative Classification and Semantic Segmentation Network (ICSSN) [116] was proposed. This network improves classification at both the object and pixel levels by iteratively enhancing the shared feature extraction module. It incorporates an object-level and sub-object-level contrastive learning method to extract crucial features, especially those at landslide boundaries. The ICSSN, tested on real-world landslide data, showed substantial improvements in classification and segmentation accuracy, outperforming baseline models in various performance metrics. Developing an effective landslide mapping system using high-resolution images is a crucial task for emergency response. However, most of existing methods rely on remote-sensing data with specific spatial resolution, making multi-scale landslide detection difficult in operational applications. To solve this, Yu et al. (2022) [117] proposed a new hierarchical deconvolution network for landslide mapping. This network enhances feature learning by using deconvolution operations and a hierarchical structure, which allows for a better synthesis of landslide features at a higher spatial resolution. Additionally, an attention module is incorporated to improve the quality of detected multiscale landslide features. In practical applications, the model was evaluated over areas in Haiti, Taiwan, and Zhouqu that recently experienced landslides, it shows significant improvements over six widely used frameworks, with a higher F1-Score (21%) and IoU (10%), demonstrating its robustness and potential for deployment in diverse geological settings for landslide mapping and management.

Another limitation for the current learning-based method in landslide mapping is a lack of sufficient training samples. To tackle this problem, Yi et al. (2020) [118] presented a novel approach using single-temporal RapidEye satellite images. It involves three main steps: the automatic generation of training samples with data preprocessing and augmentation, constructing a cascaded deep learning network named LandsNet to learn landslide features, and optimizing identified landslide maps with morphological processing. Comparative experiments in two earthquake-affected regions showed that this approach generated the best F1-Score (86.89%), which was about 7% and 8% higher than that obtained by ResUNet and DeepUNet, demonstrating its potential in emergency response and natural disaster management. To cope with the challenge of identifying old landslides, which are often visually obscured due to long-term surface processes,

Liu et al. (2023) [119] developed a feature-fusion-based segmentation network, FFS-Net, which uniquely combines texture and shape features from HRSIs and terrain features from DEM. It employs a multiscale channel attention module for feature balancing and a transposed convolution layers for image resolution restoration. Applied to the open-source dataset, FFS-Net demonstrated a superior performance compared to existing models U-Net and DeepLabV3+, showing higher metrics in the mIoU, F1 score, and pixel accuracy. A summary of case studies of landslide mapping is shown in Table 6.

**Table 6.** Summary of case studies of landslide mapping.

Study Area	Data Used	Model Evaluated	Metric and Result (%)	Reference
Huizhou City, Anhui, China	Google Earth Image	UNet, L-UNet (best)	mIoU (75.18), F1-score (85.97), Recall (83.54), Precision (88.54)	Dong et al. (2022) [109]
Zhongxinrong County, China	Sentinel-1, DEM, Slope, Curvature	UNet, SegNet, DRs-UNet (best)	IoU (93.48), F1-score (96.08), Recall (96.12), Precision (96.07)	Chen et al. (2022) [110]
Zigui County, Hubei, China	Landsat-8, Landslide Influencing Factors	ResNet50, Swin-Transformer, DeiT, DST (best)	OA (98.17), F1-score (98.16), Recall (98.16), Precision (98.16), Kappa (97.66)	Huang et al. (2023) [111]
Bijie City, Guizhou, China	TripleSat Image	ResUNet, Transformer + ResUNet (best)	mIoU (87.91), F1-score (87.73), Recall (88.23), Precision (87.24)	Yang et al. (2022) [112]
Loess Plateau, China	Optical Image	ResUNet50, SCANet (best)	OA (96.02), F1-score (90.91), Recall (90.83), Precision (90.98)	Wang et al. (2022) [113]
Bijie City, Guizhou, China	TripleSat Image	U-Net, DeepLabv3+, PSPNet + ResNet50 (best)	mIoU (91.18), Recall (96.90), Precision (93.76)	Yang et al. (2022) [114]
Yaan-Lushan, Sichuan, China	RapidEye Satellite Image	ResUNet, UNet, LandsNet (best)	OA (99.53), F1-score (77.26), Recall (78.29), Precision (76.26), Kappa (77.03)	Yi et al. (2020) [118]
Loess Plateau and Bijie Dataset	Optical Image, TripleSat Image	U-Net, DeepLabV3+, FFS-Net (best)	OA (92), mIoU (67), F1-score (59.60)	Liu et al. (2023) [119]

#### 4.4. Landslide Susceptibility Mapping

Landslide Susceptibility Mapping (LSM) is a critical step for disaster mitigation and risk management, focusing on assessing the likelihood of landslides in a region under specific geo-environmental conditions [120]. Despite its importance, the complex nature of landslide formation makes the creation of accurate and reliable LSMs a challenging task. In recent years, the rapid development of deep learning has significantly facilitated the field of LSM. Upon reviewing the latest advancements, we have found that deep learning methods in LSM can be broadly categorized into three frameworks: a classification framework, sequence framework, and hybrid framework. Each of these frameworks leverages unique technique to address specific challenges of LSM. Here, we delve into how these frameworks have been applied in LSM.

Classification Framework (CNNs, Transformer-based): Classification framework utilizes CNNs and transformer-based networks to process multiple geospatial factors and make predictions. Each factor, such as rainfall, slope, and elevation, is treated as a distinct channel in an input tensor. Here, these networks are used for extracting spatial features from inputs and identifying patterns that signify landslide-prone areas. After that, these features are typically fed into a classification head to make a prediction for landslide susceptibility map. The advantage of this framework is its capability of capturing the spatial interconnections among various factors, allowing for a comprehensive analysis of landslide susceptibility.

For example, many machine-learning and statistical methods for post-earthquake LSM often overlook the spatial structure of influencing factors. To address this, Chen et al. (2020) [121] developed a CNN model for post-earthquake LSM. The model incorporated

multiple inputs including pre-earthquake Landsat TM images and various influencing factors such as DEM, slope, and lithology. When applied to the 2008 Wenchuan earthquake area, the model demonstrated high validation accuracy (precision: 77%, recall: 90%, and F1-Score: 83%), outperforming the results of traditional logistic regression and support vector machine (SVM) models. This approach underscores the effectiveness of considering the environmental influence of a pixel and its surroundings in LSM. Jiang et al. (2023) [122] compared the performance of CNN and conventional machine learning methods in landslide susceptibility assessment. Focusing on the Hongxi River Basin area, 11 conditioning factors were considered to build the models. The study found that the CNN achieved a higher accuracy (86.41%) and AUC (0.9249) compared to six other machine-learning methods. The authors indicated that while all models successfully identified most landslides, CNN showed a stronger capability in recognizing landslide cluster regions, attributed to its convolution operation that considers environmental information.

Besides CNNs, transformer-based models are also widely explored in LSM. Distinguished by the self-attention mechanisms and ability to handle multisource inputs, transformer offers a significant advancement over traditional methods. For example, Wang et al. (2023) [123] developed a comprehensive evaluation system for landslide susceptibility using 12 key influencing factors, such as slope and elevation. They implemented a transformer model, enhanced with a self-attention mechanism and a forward propagation network, and trained it using diverse data sources. Conducted in Jiuzhaigou County, Sichuan, China, their results revealed that this transformer-based model surpassed the performance of traditional models such as random forest, SVM, and logistic regression, achieving a notable accuracy of 86.89%. This result underscores the model's enhanced capability in predicting landslide susceptibility, showcasing its superiority over conventional methods. Bao et al. (2022) [124] tackled the limitations of traditional CNNs, which predominantly focus on local information due to their fixed convolutional kernels. To address this, they proposed a novel LSM approach that integrates the Vision Transformer (ViT) and Swin Transformer, which enables a more comprehensive understanding of global spatial information. The experimental results demonstrated the Swin Transformer's superior accuracy and the ViT model's high consistency with real-world conditions, proving their strong generalization capabilities compared to traditional CNN and machine learning methods. Additionally, combining transformer with machine learning models also generated promising results for LSM. Zhou et al. (2023) [125] developed an innovative model that combines a Feature Fusion Transformer (FFTR) with various machine learning classifiers, such as Random Forests, XGBoost, and others, to enhance LSM along the Karakoram Highway in Xinjiang, China. The FFTR effectively integrated spatial information from multiple landslide conditioning factors, which were then analyzed using these classifiers. The combined FFTR-RF model exhibited outstanding performance, excelling in several key metrics like accuracy, precision, recall, and the F1-Score. This approach not only improved LSM accuracy but also provided valuable insights for developing early warning systems and landslide mitigation strategies in this region.

Sequence framework (RNN Series): the sequence framework, including RNNs [41] and their variants like LSTM [42] and GRU [43], is typically known for handling sequence and time-series data. However, in LSM, these frameworks are typically employed to process multiple geospatial factors, but not necessarily in a time-dependent manner. They focus on understanding the intricate relationships between different influencing factors, such as geological conditions, rainfall intensity, human activities, etc. By processing these factors, they can predict areas susceptible to landslides, highlighting the influence of each factor in the occurrence of landslides.

Studies by Mutlu et al. (2019), Wang et al. (2020) and Yi et al. (2022) [126–128] have demonstrated the strong capabilities of RNNs in this area. For example, Mutlu et al. (2019) focused on the assessment of landslide susceptibility in Turkey's Buyukkoy catchment area, employing RNNs to predict future landslide events based on past occurrences. Their results illustrated implemented RNN achieved a high estimation capability for landslide

susceptibility. In Yongxin County, Jiangxi, China, Wang et al. (2020) conducted a comparative analysis to the application of a standard RNN and its variants, including LSTM, GRU, and Simple Recurrent Unit (SRU) for LSM. Utilizing data from 364 historical landslide locations and 16 influencing factors, the effectiveness of these RNN-based methods was validated through various metrics. Their findings revealed that high susceptibility areas were concentrated in specific regions of Yongxin County, and all RNN models demonstrated strong predictive performance, suggesting their applicability in landslide risk management. On the Qingchuan County, Sichuan, China, Yi et al. (2022) evaluated the effectiveness of three classical neural networks: an Artificial Neural Network (ANN), 1D CNN, and RNN. This research highlighted the superior performance of the RNN in both qualitative and quantitative aspects, in comparison to other models, which suggests the potential of RNNs in enhancing landslide susceptibility predictions.

In the Kush–Himalayan region along the China–Nepal Highway, Xiao et al. (2018) [129] adopted data-driven algorithms for LSM. Incorporating various factors such as elevation, vegetation, and precipitation, they evaluated the result of four machine learning models: Decision Tree (DT), SVM, ANN, and LSTM. The study demonstrated that LSTM excels notably due to its ability to effectively manage time series data with extended temporal dependencies. This highlights the importance of considering dynamic geological and environmental factors in landslide risk assessments. Another study conducted by Ji et al. (2023) [130] indicated that the selection of non-landslide samples plays a significant role in the LSM using deep learning methods. In the research within Xinhui District, China, they utilized information value analysis to refine sample selection for RNN models, including the Simple Recurrent Unit (SRU) and standard RNN. Their methodology, which focused on the strategic prioritization of factors influencing landslides, significantly enhanced the predictive accuracy of the models. Among four tested models, RNN and SRU with and without optimized sample selection, they found that the RNN with optimized selection emerged as the most effective method, achieving superior results in the AUC and other key metrics, which underscores the importance of careful sample selection in enhancing LSM model performance.

**Hybrid Framework (spatial temporal):** The hybrid approach combines the advantages of spatial and temporal networks. In this framework, spatial network, like CNNs, are first used to process the input data, extracting relevant features from various geospatial factors. These extracted features are then fed into the temporal network, like RNNs, for analyzing the relationships and dependencies between these features to predict landslide susceptibility. This combination harnesses the spatial feature extraction strength of CNNs and the sequence modeling prowess of RNNs, making the hybrid framework particularly effective in handling complex LSM tasks.

The simplest configuration of a hybrid model is two basic units' combination (CNN and RNN). For instance, Li et al. (2021) [131] presented a hybrid framework for LSM in the Three Gorges Reservoir (TGR) area. Specifically, they created a spatial database incorporating 20 landslide conditioning factors and 196 landslide polygons, and constructed an ensemble model integrated with CNN and RNN. Applying this framework into practice, this framework presented superior predictive capability with various metrics compared to standalone CNN, RNN, and logistic regression models in landslide disaster management. Traditional machine learning methods for LSM often rely on fully connected layers to select influence factors, which can limit feature extraction efficiency. To overcome this, a novel CNN-LSTM [132] model with an attention mechanism was proposed. This model aims to avoid the complex optimization of input factors while achieving the same or even better prediction accuracy. Validated with historical landslide data from Kerala, India, the experimental results demonstrated that proposed model significantly outperforms existing methods like independent CNN and LSTM in terms of prediction accuracy.

In addition to regular geospatial factors, Yuan et al. (2022) [133] indicates that incorporating InSAR data can benefit landslide susceptibility analysis thanks to its capability of mapping surface deformation. In detail, their approach comprises two key stages:

firstly, it extracts landslide predisposing factors from various sources including InSAR data; secondly, it constructs a hybrid model combining CNN with RNN variants, including CNN-LSTM, CNN-GRU, and CNN-SRU. Evaluated using various metrics, this method achieved a higher performance than other CNN-Machine Learning based models, and the results were more precise than when using the same methods without considering InSAR deformation features. To tackle the challenge of comprehensive landslide hazard assessment on a national scale, this study [134] developed a novel deep learning method for LSM. Specifically, this approach integrated three hybrid networks (CNN-SRU, CNN-LSTM, and CNN-GRU) for landslide susceptibility analysis, and a spatiotemporal transformer for temporal probability prediction. Subsequently, to address geographical regional differences, the results were further refined using a landslide hazard formula that employs the FR method to calculate weights for landslide spatiotemporal probability in different areas. Validated across the United States, this method shows excellent performance compared with existing works, which proves its practical significance for national-scale landslide hazard assessment. Viewing landslides as multi-dimensional objects, Wang et al. (2021) [135] introduced an AI-powered, object-based method for performing LSM. This approach uses historical data to define landslide objects and integrates geo-environmental data to organize samples. Evaluated with six algorithms in a Hong Kong, experimental results showed that BiLSTM-RNN and CNN-LSTM presented superior performance compared with the other AI models. This object-based approach significantly outperformed traditional methods and produced Hong Kong's first AI-based, territory-wide landslide susceptibility map. A summary of case studies of landslide susceptibility mapping is shown in Table 7.

**Table 7.** Summary of case studies of landslide susceptibility mapping.

Study Area	Data Used	Model Evaluated	Metric and Result (%)	Reference
Wenchuan County, Sichuan, China	Landsat Image, DEM, Slope, and Lithology	LR, SVM, CNN (best)	F1-score (83), Recall (90), Precision (77)	Chen et al. (2020) [121]
Pingwu County, Sichuan, China	Lithology, Elevation, Slope, Aspect, Roughness, NDVI, Curvature, and Land Cover	MLP, RF, DT, GBDT, Adaboost, Naive-Bayes, CNN (best)	Accuracy (86.41), AUC (92.49)	Jiang et al. (2023) [122]
Jiuzhaigou County, Sichuan, China	Elevation, Slope, Lithology, Seismic Intensity, Land Use, and Annual Rainfall, etc.	RF, SVM, LR, Transformer (best)	Accuracy (86.89), AUC (91.50)	Wang et al. (2023) [123]
China-Nepal Highway, Kush-Himalayan, China	DEM, Geologic Maps, High Resolution Optical Image, and Meteorological Data	BPNN, SVM, DT, LSTM (best)	Accuracy: BPNN (62), SVM (72.9), DT (60.4), LSTM (81.2)	Xiao et al. (2018) [129]
Three Gorges Reservoir (TGR), China	Altitude, Aspect, Slope, Curvature, Land Use, NDVI, Rainfall and Lithology, etc.	CNN, RNN, CNN + RNN (best)	OA (86), F1-score (86.98), Recall (93.62), Precision (81.22)	Li et al. (2021) [131]
Shuicheng County, Guizhou, China	InSAR, Topography, Geomorphology, Geology, and Hydrology, etc.	CNN-LSTM, CNN-SVM, CNN-SRU, CNN-GRU (best), CNN-RF, CNN-LR	AUC (98.40), ACC (93.70), MCC (87.50), Kappa (87.40)	Yuan et al. (2022) [133]

#### 4.5. Landslide Displacement Prediction

Landslide displacement prediction is an essential aspect of geohazard analysis focused on forecasting the movement and behavior of landslides [136]. This task is challenging due to the complex and dynamic nature of landslide phenomena. However, the deep learning method, with its ability to analyze complex and nonlinear patterns, provides a promising solution. In this field, two primary deep learning frameworks are utilized: the sequence framework and the hybrid framework. The sequence framework excels in capturing temporal dynamics of landslide movements, while the hybrid framework integrates both spatial features and temporal patterns for a comprehensive analysis. In the following

section, we will delve into recent advancements in these frameworks, highlighting how they contribute to accurate and efficient landslide displacement predictions.

**Sequence framework (RNN Series):** This framework utilizes sequence networks, such as RNN, LSTM, and GRU, to model the evolution of landslides and predict their future movement trend. These networks are trained on historical landslide data, including various geospatial and environmental factors. These factors often include rainfall intensity, geological conditions, slope stability data, and other relevant variables. The strength of this framework lies in its ability to model the temporal dynamics and nonlinear relationships inherent in these factors, providing a robust mechanism for accurate prediction of future landslide activities.

Among various models, RNNs stand out for their simplicity and efficacy in predicting sequence data, making them suitable for forecasting landslide deformation. Recently, Chen et al. (2015) [137] utilized a RNN to predict Hubei Baishuihe landslide displacement. Adopting the Levenberg–Marquardt algorithm and a genetic algorithm for model training, the experiments demonstrated that RNN significantly outperforms the feedforward neural network with a higher prediction accuracy, which establishes the RNN's effectiveness and feasibility for landslide displacement prediction. To model the complex, nonlinear component of landslide displacements, Zhang et al. (2022) [138] employed a GRU to analyze the Jiuxianping landslide in China. The GRU's dynamic properties enabled it to effectively capture periodic variations in displacement, offering a more accurate prediction than static models such as an Artificial Neural Network, random forest, and multivariate adaptive regression splines (MARS). The success of the GRU model in this context demonstrates its potential as a valuable tool for early warning systems for landslide displacement prediction. To find out the suitable model for landslide deformation prediction, Xi et al. (2023) [139] conducted a comprehensive comparison of eleven machine learning models, focusing on their ability to predict slope deformation using time-series data from the Huanglianshu landslide. The study highlighted the superiority of the LSTM model, particularly when enhanced with an attention mechanism and a transformer module. This configuration delivered the highest accuracy in predictions, underscoring the transformer module's effectiveness in dealing with nonlinear landslide displacements influenced by a variety of factors. The findings provide valuable insights into selecting the most suitable models for early landslide warning systems.

Contemporary research mainly focuses on static models for landslide displacement prediction. However, influenced by geological and environmental factors, the dynamic, nonlinear, and unstable nature of landslide deformation poses significant challenges in the displacement prediction. To enhance the accuracy of prediction, a prevalent approach involves decomposing the displacement into trend and periodic components, and modeling each component independently. This method allows for a nuanced capture of landslide characteristics by separating these components. Typically, statistical and empirical models are utilized to assess the trend component, while deep learning networks are applied for periodic component modeling. For example, in their respective studies on landslide displacement prediction, Yang et al. (2019), Zhang et al. (2021), and Xie et al. (2019). [140–142] each adopted dynamic models that specifically focus on decomposing landslide displacement into trend and periodic components. The approach of Yang et al. (2019) in the TGR region employed time series analysis and LSTM, effectively modeling the periodic displacement influenced by environmental factors. The LSTM model was validated on two TGR landslides, showing superior performance over the static SVM model. Zhang et al. (2021) also targeted the TGR region with a method integrating GRU and time series analysis, where the trend component was modeled using a cubic polynomial and the periodic component with GRU. This method significantly enhanced prediction accuracy compared to static models. Similarly, Xie et al. (2019) approached Laowuji landslide prediction by incorporating factors like geological conditions into their model, using empirical mode decomposition for the trend and LSTM for the periodic component, which highlights LSTM's capability in capturing landslide dynamics. To sum up, these studies collectively emphasize

the importance of dynamic modeling in accurately predicting landslide displacements in complex and variable environments.

**Hybrid framework (spatial temporal):** The hybrid framework merges the spatial analysis capabilities of CNNs with the temporal data processing strengths of RNNs, making it highly effective for landslide displacement prediction. In this approach, spatial networks extract and analyze spatial features from time series grid data, identifying key patterns in factors such as geological conditions and environmental changes. These extracted features are then input into temporal networks, like RNN, LSTM and GRUs, which are adept at understanding temporal sequences and trends. This integration allows for a thorough analysis of both the spatial and temporal aspects of landslide dynamics, leading to more accurate and comprehensive displacement predictions.

When modeling displacements, traditional models have always overlooked the aspect of random displacement. To solve this problem, Lin et al. (2023) [143] introduced a hybrid CNN-BiLSTM model to enhance the displacement prediction. Specifically, they proposed a Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN) method for the landslide displacement decomposition, and adopted the Grey Relational Analysis–Maximal Information Coefficient (GRA-MIC) calculation for selecting key factors. This hybrid model is adept at capturing both the temporal and spatial characteristics of data, thereby enhancing the accuracy of landslide displacement prediction. In the Baishuihe landslide area within the TGR Area, the author implemented the CNN-BiLSTM model to predict landslide displacement. This area, with its complex terrain and substantial landslide risk, provided a critical testing ground for the model's capabilities. Comparative analysis with seven models under four distinct input scenarios shows that the CNN-BiLSTM model is more effective than the single model in predicting landslide displacement, which holds potential for early landslide warning, benefiting the geological and land use policy decision-making process in such high-risk areas.

Currently, more and more deep learning models have been applied in landslide displacement prediction; thus, a comprehensive comparison to their effect is urgently needed. Nava et al. (2023) [144] evaluated seven deep learning methods for their effectiveness in forecasting landslide displacement, focusing on various landslides with different geographical contexts. The models include multi-layer perception (MLP), LSTM, GRU, 1D CNN, 2xLSTM, bidirectional LSTM, and a hybrid model Conv-LSTM. The results revealed that different models have varying effectiveness rates in predicting landslide displacement. The MLP, GRU, and LSTM models consistently provided reliable predictions across different landslide scenarios. Specifically, the Conv-LSTM model showed superior performance in predicting the highly seasonal Baishuihe landslide. Interestingly, the study highlighted that the MLP model excelled in forecasting higher displacement peaks, while LSTM and GRU were more adept at modeling lower displacement peaks. These findings suggest the importance of selecting the appropriate deep learning model based on the specific characteristics of the landslide. Considering the dynamic, nonlinear, and unstable nature of landslides, Li et al. (2023) [145] proposed a dynamic model based on CNN-LSTM for landslide displacement prediction. This model decomposes displacements into trend, periodic, and random components, with a least square quintic polynomial function modeling the trend and CNN-LSTM predicting the periodic and random displacements. Tested on the Bazimen landslide in TGR, the model showed higher accuracy and stability than the LSTM, GRU, and neural networks, significantly reducing errors in periodic displacement prediction.

While temporal dependencies in landslide displacement have been extensively studied, the exploration of spatial dependencies remains limited due to the significant variations in the spatial structures of landslides. To address this, a novel Graph Convolutional GRU Network (GC-GRU-N) [146] has been developed. The model employs graph convolutions with weighted adjacency matrices to capture spatial dependencies in landslide structures; then, these features were fed to GRU for temporal dependency learning. Applied to two landslides in the TGR, GC-GRU-N demonstrated superior performance over comparative models. The inclusion of external factors in attribute augmentation notably enhanced the

prediction accuracy. In a similar approach, Kuang et al. (2022) [147] introduced a novel landslide prediction model employing graph neural networks to tackle the challenge of capturing complex spatial deformations and their interdependencies across various areas. The model integrates graph convolutions, which aggregate spatial correlations among different monitoring sites, and a locally historical transformer network for dynamic spatial-temporal analysis. Through extensive real-world data experiments, the model demonstrates superior performance over existing methods that solely rely on time-series data, both in prediction accuracy and interpretability. To analyze the complex interactions of spatial and temporal factors in landslide displacement, Yang et al. (2023) [148] introduced an innovative model that effectively combines a CNN and a LSTM network. The model employs the Maximal Information Coefficient (MIC) for analyzing spatial-temporal correlations in GNSS data. Enhanced with spatial-temporal attention mechanisms, this model was applied to the Outang landslide in the TGR Area. The results indicate that the proposed model is highly effective in predicting displacements consistently across all six monitoring points. In contrast, other models, while showing strong performance at certain individual stations, do not maintain this level of accuracy on all the locations. A summary of case studies of landslide displacement prediction is shown in Table 8.

**Table 8.** Summary of case studies of landslide displacement prediction.

Study Area	Data Used	Model Evaluated	Metric and Result	Reference
Jiuxianping Landslide, Yunyang County, China	Historical Displacement, Rainfall, Reservoir Level	ANN, RF, MARS, GRU (best)	MAPE (0.002), RMSE (2.169), $R^2$ (99.90%), Bias Factor (1.001)	Zhang et al. (2022) [138]
Baishuihe landslide, TGR, China	Historical Displacement, Rainfall, Reservoir Level	CNN-BiLSTM (best), CNN-RNN, CNN-LSTM, CNN-GRU, etc.	MAE (1.789), RMSE (2.206), MAPE (0.078), $R^2$ (99.84%)	Lin et al. (2023) [143]
Shuping landslide, TGR, China	Historical Displacement, Rainfall, Reservoir Level	GC-GRU-N (best), T-GCN, LSTM, MLR, SVR, etc.	MAE (6.123), MASE (0.353), RMSE (8.321),	Jiang et al. (2022) [146]
Houziyan Dam, Danba County, Sichuan, China	Historical Displacement	LandGNN (best), LSTM, GRU, SVR, DCRNN, etc.	MAE (0.106), RMSE (0.132), ACC (0.892), $R^2$ (34.80%)	Kuang et al. (2022) [147]
Outang landslide, TGR, China	Historical Displacement, Rainfall, Reservoir Level	LSTM, CNN-LSTM, Attention CNN-LSTM (best)	MAE (0.99), RMSE (1.18), MAPE (0.33), $R^2$ (99.89%)	Yang et al. (2023) [148]

## 5. Challenges and Opportunities

This section includes the challenges and emerging opportunities of utilizing deep learning in landslide studies mostly with remote sensing imagery. We will discuss the main obstacles that currently limit progress and consider how they might be addressed. Additionally, we will spotlight promising research areas within deep learning that show potential in addressing landslide scientific and engineering problems.

### 5.1. Challenges

#### 5.1.1. Label Acquisition

A significant challenge of applying deep learning methods for landslides is how to acquire a large number of high-quality labels. This challenge is a key obstacle across various tasks, as good and sufficient labels are fundamental for the effective training of all the supervised learning models. For landslide detection and mapping, labels mostly come from manually labeled data from remote sensing images, on which experts annotate landslide features or areas. For LSM, labels often originate from a combination of historical landslide inventory data, terrain attributes, and land cover data, which are used to identify areas prone to landslides. As for landslide displacement prediction, labels can indeed come from various sources, including time-series analysis of satellite imagery, and ground-based

monitoring data such as GPS or inclinometer measurements, as well as historical records of landslide movement or deformation captured through field surveys.

Given the complex nature and non-linear dynamic of landslides, it still lacks a well-organized label database for landslide studies. The historical landslide records are often faced with issues of completeness and accuracy, particularly in remote or less monitored regions where comprehensive data collection is challenging, leading to significant gaps in records [149]. Additionally, a considerable portion of these records are derived from manual labeling on remote-sensing images facilitated by field surveys. This manual process, while crucial, is time-intensive and geographically limited, typically confined to areas that are easily accessible [150]. This limitation in data collection poses a challenge in obtaining a thorough and accurate dataset for landslide analysis.

Furthermore, the dynamic nature of environmental factors also poses a challenge to label's accuracy. The susceptibility of certain areas to landslides can rapidly change due to environmental fluctuations [151]. Therefore, in the process of training models, particularly when utilizing multi-source data, researchers must be vigilant about the time of data acquisition. This ensures that the labels used are consistent with the conditions at the time of data collection. In addition, landslide displacement prediction relies heavily on precise, continuous historical data, typically sourced from advanced techniques like GPS or InSAR [152]. However, the cost and technological demands of these methods can be substantial. For example, monitoring slow-moving landslides requires continuous data collection over a decadal time scale, which can be challenging to sustain.

Finally, the process of acquiring high-quality labels for landslide analysis necessitates an interdisciplinary approach, integrating knowledge from geology, hydrology, and meteorology. This requirement adds more complexity, as understanding phenomena like the hydrological triggers of landslides extends beyond standard geological expertise. Moreover, data privacy and accessibility issues also cause challenges, in which restrictions on accessing detailed land surveys or satellite imagery prevents the availability of crucial information needed for comprehensive analysis.

These factors collectively underscore the complexities and resource demands with label acquisition in landslide analysis using supervised learning, emphasizing the need for cross-disciplinary collaboration and advanced data gathering techniques.

### 5.1.2. Model Generalizability

Model generalizability poses another challenge in applying deep learning techniques to landslide studies. This difficulty arises primarily from two aspects: data-related challenges and model training issues. These factors contribute to the complexity of ensuring that a trained model performs effectively in real-world scenarios.

Data-related challenges: Externally, the variability in geographical and environmental conditions, along with changing climate and land use patterns, result in the complexity of observations [153]. Each region has its unique characteristics influencing landslide behavior, making it challenging for a model trained on one region to perform accurately in another region. This complexity comes from spatial and temporal heterogeneity. Landslides are influenced by a various of factors that change over time and vary across different locations. For example, a model trained on data from a region experiencing frequent heavy rainfall might not predict landslide displacement accurately in an area with a drier climate. This heterogeneity requires models to be adaptable to a wide range of environmental conditions, or only applicable regionally. Additionally, scalability to different areas is also a challenge. Models that are effective in small, controlled study areas often face difficulties when scaled up to larger regions [154]. This is because larger areas exhibit a greater variety of geological and climatic conditions, which complicates the task for a single model to comprehensively capture and understand all these diverse nuances.

To mitigate these challenges, incorporating multi-source data can be highly beneficial [155]. By integrating diverse data sources such as geological surveys, and hydrological and land-use data, the model gains a more comprehensive view of the influencing factors.

These broader data help the model adapt to different environmental conditions and improve its predictive accuracy across various regions. For the scalability problem, it can be mitigated by implementing some scalable machine learning algorithms. These algorithms were designed to adjust their complexity in accordance with the geographical scope of the data [156]. Such approaches improve the performance of a model in both small- and large-scale regions.

**Model training issues:** Overfitting is a common issue of deep learning methods, which makes the model fail to generalize to new, unseen data [19]. This issue manifests in two primary ways. Firstly, overfitting can occur when the model is over-tuned to the nuances of the training data, making it less effective at handling new data. This is often a technical issue that can be mitigated through regularization, where the model is tested on various subsets of the data to ensure its ability to generalize.

Secondly, overfitting arises when the training data are not diverse enough, leading the model to learn patterns that are not universally applicable. To address this, domain adaptation techniques can be employed. These techniques involve adjusting the model to perform better in a new domain that it was not originally trained on, enhancing its adaptability to different scenarios. Additionally, ensemble learning can be another powerful strategy to improve generalizability [157]. This approach involves combining multiple models, each potentially excelling in different aspects of data interpretation.

In summary, the challenges of model generalizability are multifaceted for applying deep learning for landslide studies. While data-related challenges, such as spatial, temporal heterogeneity and scalability, may be complex to address, strategies like incorporating multi-source data, domain adaptation, and ensemble learning offer effective ways to enhance the adaptability and accuracy of these models.

### 5.1.3. Multi-Source Data Integration

How to intergrade multi-source data poses a challenge to landslide studies, especially when leveraging learning-based techniques. In this task, achieving reliable predictions requires researchers to adopt data from various sources to train their models. These include satellite imagery, aerial photography, ground-based sensors, historical landslide records, geological surveys, and meteorological data [158], etc. Each data source offers unique insights, such as terrain stability, weather patterns, and historical landslide occurrences, etc. However, processing these varied data to make them suitable for model training is a complex task.

The first challenge is handling different data format and structure. Data from different sources often come in various formats and structures. Integrating them into a format suitable for deep learning models requires a deep understanding of related fields. For example, radar remote sensing data record the amplitude and phases of microwaves reflected from the surface, processing and understanding them require specialized knowledge in radar signal. Similarly, geological surveys might provide data in textual reports or tabular formats, which necessitates different processing techniques. Therefore, a mature and versatile data pre-processing system is urgently needed to help standardize and streamline this process.

Another challenge is ensuring spatial and temporal alignment. This involves aligning data from different sources to the same geographical locations and time frames [159]. For example, consider integrating historical landslide records with recent satellite imagery. These historical records provide valuable insights into past landslide occurrences, but aligning them with current satellite images can be complex due to changes in landscape over time and differences in geographical coordinates. Temporal alignment adds another layer of complexity. Satellite imagery captures land conditions on the acquisition dates, but aligning this with events from the past, which may have occurred in a different landscape. To address these challenges, techniques such as georeferencing for accurately mapping historical data onto satellite images, and advanced algorithms for data interpolation to reconcile time frame differences, are essential [160]. These techniques ensure that the integrated data accurately reflect both past and current conditions for effective landslide prediction.

Finally, developing a practical and efficient data fusion algorithm remains difficult. Most current approaches treat data from various sources as individual channels within a high-dimensional tensor, then feeding it into the model as a whole. For instance, meteorological data, such as rainfall intensity, and geological data, like soil type, differ significantly in both their nature and scale. Merely stacking these diverse data as layers in a tensor may hinder a deep learning model to effectively discern the intricate relationships among these variables. Thus, a more reasonable strategy might entail developing model architectures with specialized components, each specifically designed to handle the unique attributes of individual data sources. This tailored approach could significantly enhance the model's ability to interpret and learn from multi-source data.

#### 5.1.4. Model Interpretability

The interpretability of the deep learning models is not just an academic concern but a practical necessity for landslide studies. Given the high-risk nature of landslides, decisions can have significant safety and economic implications, gaining insights into the interpretability of a model and understanding how it arrives at its predictions is vital. The challenge in model interpretability arises primarily from the inherent complexity of deep learning algorithms. These models process information in ways that are not always intuitive to humans, making it difficult to trace how input data lead to specific predictions [161]. In practical applications, the rationale behind a model's prediction is as crucial as the prediction itself. For example, if a deep learning model identifies a high risk of landslide in a specific area, decision-makers must understand the basis of this prediction. Is it due to recent rainfall patterns, geological instability, or a combination of both factors? Such deep insight is invaluable for informed decision-making, ensuring that responses are not only swift but also precisely targeted to address the specific contributing factors.

To address these challenges, several methods are being developed and employed. Visualization techniques that expose layers and activations within CNNs can offer insights into what the network is focusing on [162]. For instance, identifying those certain filters in the network are particularly responsive to changes in topography can be crucial for understanding model focus areas. These tools bridge the gap between complex model outputs and feasible strategies.

In real-world scenarios, interpretability tools like SHapley Additive exPlanations (SHAP) have been employed to clarify model predictions [163]. In landslide susceptibility mapping, SHAP values can indicate the importance of various input features, enhancing the understanding of model outputs. Beyond SHAP, interpretability in deep learning models can be enhanced by employing a variety of techniques such as Local Interpretable Model-Agnostic Explanations (LIME) [164], Class Activation Mapping (CAM) [165], and Gradient-weighted Class Activation Mapping (Grad-CAM) [166]. LIME aids interpretability by approximating the model locally to highlight the contribution of each feature to the prediction, proving useful to understand the model's decisions for specific instances. CAM offers a visualization method to discern which regions of the input significantly impact predictions in CNNs. By generating a heatmap of class activations, researchers can visualize which features are most influential in classification results. Furthering this visualization technique, Grad-CAM utilizes the gradient information flowing into the final convolutional layer to create a localization map that indicates critical image regions for predictions. This approach is especially helpful for providing visual explanations of CNN decision-making processes in landslide risk assessment.

The applicability of these methods in the landslide studies has been substantiated by several examples. A study [167] utilizing the SHAP-XGBoost framework demonstrated how SHAP values can elucidate the influence of various regional factors on landslide susceptibility, highlighting the spatial heterogeneity in the contributions of different factors. This analytical capability is integral for crafting landslide mitigation strategies. In another instance [168], deep learning models like Resnet-50 and VGG-19, coupled with Grad-CAM and Score-CAM, have been effectively employed for landslide detection in Turkey's Gün-

doğdu area. The visualization techniques offered by Grad-CAM and Score-CAM provided crucial insights into the imaged regions most indicative of landslides, confirming the models' accuracy and enhancing the trustworthiness of the detection systems. Furthermore, a novel method [169] combining CAM with a cycle generative adversarial network (cycleGAN) was proposed to overcome the challenge of extensive data labeling for landslide segmentation in remote sensing images. Using image-level annotations, CAM identified approximate landslide zones, which were then refined by cycleGAN, significantly reducing manual annotation work and yielding better mIoU scores, thus validating the potential of weakly supervised learning in this application.

In addition, attention mechanisms in deep learning offer another approach for increasing interpretability [170]. In a landslide prediction model, these mechanisms could highlight regions in an input image that most influences the model's prediction. This feature not only aids in model interpretability but also assists in validating the accuracy and relevance of the model focus. The balance between model complexity and interpretability is a critical consideration. More complex models might offer higher accuracy, but their opaque nature can be a significant drawback for interpretation. In some cases, a slightly less accurate but more interpretable model, such as a decision tree, might be preferable. These simpler models can sometimes be used in conjunction with neural networks, providing both high-accuracy predictions and a more understandable decision path [171]. Ultimately, the goal in leveraging deep learning for landslides is to create models that are not only accurate but also transparent and understandable.

#### 5.1.5. Computational Demands

The utilization of computational resources for training deep learning models in landslide research is influenced by several key factors. The complexity of the model, often quantified by Floating Point Operations (FLOPs), is a primary determinant, because models with more parameters require computational power. Additionally, the scale of the dataset and the difficulty of training—reflected in the model's convergence behavior and the time it takes to train—significantly impact the number of computational resources needed.

Training resource-intensive models, such as large language models (LLMs) like GPT or BERT, demands a substantial investment in computational power, typically necessitating industry-scale hardware clusters rather than the resources commonly available in university labs. For instance, training an LLM can consume thousands of GPU days, requiring considerable energy usage and financial cost. However, training such models from scratch is often impractical for individual research groups due to the exorbitant costs and computational requirements.

In light of these constraints, future research directions point towards two viable approaches. One practical strategy is to fine-tune publicly available pre-trained LLMs to specific tasks within the landslide domain. This method avoids the extensive computational effort of training a model from scratch, saving resources while maintaining the model's advanced features. By fine-tuning locally custom datasets, researchers can adapt these powerful models to their specific needs without the prohibitive expense of full model training.

Another direction is to explore a wider array of AI techniques beyond LLMs. Traditional deep learning models, although have much fewer parameters, sometimes can be entirely adequate for certain landslide-related tasks. These models can often be trained effectively using local GPU resources, which are more accessible to research institutions. For example, a YOLO model [99] trained on a dataset containing thousands of annotated satellite images of landslide occurrences can effectively identify vulnerable areas with high precision, demonstrating the feasibility of employing more resource-efficient models in landslide research. By embracing these adaptive strategies, researchers can effectively navigate the computational challenges posed by advanced AI models.

## 5.2. Opportunities

### 5.2.1. Physical Informed Neural Network for Reliable Modeling

The integration of Physics-Informed Neural Networks (PINNs) into deep learning for understanding landslide mechanisms is an emerging area that promises to revolutionize how we approach this complex challenge. PINNs are particularly compelling because they blend the rigor of physical laws with the adaptability of machine learning, creating models that are not only data-informed but also physically grounded [172].

The greatest potential of PINN is their ability to incorporate established physical principles directly into neural network architectures. This integration is crucial, especially in a field like landslide displacement prediction, where understanding the underlying physical processes is as important as the data themselves. In this situation, a PINN designed for landslide prediction could inherently understand slope stability principles, ensuring that its predictions are aligned with geotechnical realities. In addition, this approach significantly enhances the accuracy and reliability of predictions. Traditional machine learning models, reliant solely on data patterns, can occasionally produce results that are physically implausible. Meanwhile, PINNs may ensure that predictions adhere to fundamental laws of physics.

Another critical advantage of PINNs is their efficiency in limited observations. Landslide-related data can often be limited or inconsistent, particularly in less accessible regions. PINNs can offset this limitation by leveraging physical laws to make informed predictions, even in the face of sparse data. This ability makes them especially useful in regions where traditional data collection is challenging or where historical data are lacking. Moreover, the generalizability of PINNs across different geological and environmental conditions is a significant asset. The physical laws that PINNs integrate are universal, allowing these models to adapt to new areas with minimal retraining. This generalizability is valuable for landslide prediction efforts, which often need to be tailored to diverse and changing landscapes.

Looking ahead, the future research of PINNs for landslides is rich with possibilities. Key among these is enhancing the integration of complex geophysical dynamics into PINNs. This involves embedding detailed representations of landslide processes, such as soil erosion and water infiltration, directly into neural network models. Such advancements would enable a more precise simulation of landslide triggers and behaviors. Additionally, adapting PINNs to account for the impacts of climate change also presents a vital research avenue. Integrating climate models into PINNs could provide insights into how shifting weather patterns and extreme events might influence landslide occurrences, crucial for long-term mitigation and planning.

### 5.2.2. Large Pretrained Models for Landslides

In deep learning, the remarkable success of large models in fields like computer vision (CV) and natural language processing (NLP) has revealed an astonishing development. Models such as the GPT [82] and BERT [79] demonstrate a unique phenomenon known as “emergent behavior”. This is characterized by their ability to identify complex patterns and gain insights from massive datasets, which extend well beyond the scope of their initial programming. Emergent behavior arises from the models’ extensive and diverse training, enabling them to learn and apply knowledge in unforeseen or novel contexts. It is this capacity that has made them incredibly effective in fields like CV and NLP, where they can understand nuances in language or identify intricate patterns in images that were not explicitly included in their training.

Translating this concept to the field of landslide studies, there is a significant opportunity waiting to be explored. Data related to landslides are rich, encompassing diverse types such as satellite imagery, historical weather patterns, geological data (pointwise and maps) and text reports [173]. However, the application of large and advanced machine learning models in this area is still untapped. The potential applications of these models can be various. For instance, a large model trained on extensive landslide-related datasets

could identify subtle correlations and predictors, potentially unveiling new insights into landslide triggers and susceptibilities. This approach could lead to more accurate and nuanced predictions, factoring in a broad spectrum of environmental and climatic variables.

Additionally, developing hybrid large models for spatial and temporal data analysis also represents a promising avenue. Such a model, proficient in both processing and learning from these distinct data types, could provide a holistic understanding of landslide risks. This approach would effectively merge real-time environmental monitoring with the analysis of historical data, offering a more comprehensive perspective on potential landslide threats. Furthermore, landslide research currently relies mainly on traditional sequence models like RNNs, LSTMs, or GRUs for temporal data analysis. However, the opportunity to utilize more advanced, large pretrained models, similar to GPT or BERT, has yet to be fully explored. These sophisticated models have the potential to bring a new level of depth to the analysis of temporal data, such as weather patterns, time-series displacement and historical landslide incidents, potentially surpassing the capabilities of existing models.

### 5.2.3. Contrastive Learning for Advanced Data Analysis

Contrastive learning (CL), an advanced technique in unsupervised learning, it captures key features of datasets by contrasting positive (similar) pairs against negative (dissimilar) pairs, making it ideal for exploring large, unlabeled datasets common in environment studies [174].

A critical application of CL lies in the development of pre-trained large models for landslide-related tasks. Traditionally, the bottleneck in creating such models has been the scarcity of labeled data. CL, however, circumvents this by effectively utilizing the vast quantities of available unlabeled data. By training on these extensive datasets, a CL-based model can develop a deep understanding of the features and patterns inherent to landslide data. This pre-training phase equips the model with a comprehensive knowledge base, which can then be fine-tuned for specific tasks like landslide mapping, detection, etc. CL-based models save considerable time in downstream task training and enhances the performance thanks to the advanced initial understanding.

In landslide studies, the integration of diverse data types is essential for accurate analysis and prediction [175]. CL can be applied to enhance this multi-source data fusion process. Through its distinct methodology, CL adeptly contrasts features from various data sources, enabling it to effectively pinpoint crucial information within each dataset. This approach is instrumental in synthesizing a comprehensive view from disparate data types, such as remote sensing imagery, geological surveys, and meteorological data. For example, consider the fusion of satellite imagery with geological survey data. Satellite imagery provides a broad, top-down view of a region, capturing details about vegetation, terrain structure, and surface moisture. Geological data, on the other hand, offers in-depth insights into the soil composition, rock types, and underground structures of the same region. CL can be applied to contrast these different datasets, identifying areas where the information overlaps or complements each other. This might involve distinguishing terrain features visible in satellite imagery that correlate with geological findings, such as identifying areas of loose soil or potential fault lines that are prone to landslides.

Furthermore, CL is also suitable for analyzing homogeneous data sources over time, such as time-series satellite images, to detect landscape changes [176]. This approach is critical in change detection, where CL models contrast images from different periods to identify terrain alterations indicative of landslides. For example, a CL model might analyze two sets of satellite images of the same region taken years apart. By contrasting these images, it can pinpoint areas where the landscape has changed, such as shifts in soil patterns or vegetation cover, which could signal a potential landslide. This method is particularly valuable for early warning systems. It enables the detection of subtle changes in the landscape that precede landslides, like minor shifts in terrain, providing crucial information for preemptive measures. Additionally, in a long-term context, this

application of CL allows for the continuous monitoring of a region, contributing to a deeper understanding of how landscapes evolve over time and how environmental factors influence these changes.

## 6. Conclusions

In this work, we conducted a comprehensive review of deep learning frameworks applied in landslide studies. Initially, we provided an in-depth introduction to fundamental deep learning concepts, covering methods, models, loss functions, evaluation metrics, and architectural modules, to establish the basic knowledge of deep learning. Following this, we summarized five key deep learning frameworks, encompassing three spatial frameworks (classification, detection, and segmentation), one temporal framework (sequence), and one hybrid spatial–temporal framework, which collectively address a broad range of tasks in landslides studies.

We then showcased the latest landslide studies applying these frameworks, reviewing four critical areas: landslide detection, mapping, susceptibility mapping, and displacement prediction. Each task was discussed in relation to the corresponding frameworks. This categorization aids readers in understanding the most suitable frameworks for various tasks, enhancing their grasp of applications applying these frameworks.

Our analysis revealed that deep learning has been extensively applied in landslide studies, especially in detection tasks using two-stage and one-stage networks. However, the exploration of advanced networks like transformer-based DERT remains uncharted. In landslide mapping, segmentation networks have successfully achieved precise, pixel-level landslide mapping. We observed that in landslide susceptibility mapping, sequence and hybrid frameworks outperform traditional classification frameworks, effectively indicating high risk areas. Similarly, for predicting landslide displacement, sequence and hybrid frameworks have demonstrated superior capabilities in capturing historical data and making accurate predictions, surpassing conventional methods.

Finally, we discussed the challenges and future opportunities in this field. Key challenges include label acquisition, model generalizability, multi-source data fusion, and model interpretability. Looking ahead, we identified promising research areas such as physically informed neural networks, large pretrained models, and contrastive learning, each holding great potential for future exploration.

**Author Contributions:** Conceptualization, Q.Z. and T.W.; writing—original draft preparation, Q.Z.; writing—review and editing, T.W.; supervision, T.W.; funding acquisition, T.W. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Ministry of Science and Technology of China, under award number 2021YFC3000403 and the National Natural Science Foundation of China, under award number 42374019.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Flentje, P.; Chowdhury, R. Resilience and sustainability in the management of landslides. In *Proceedings of the Institution of Civil Engineers-Engineering Sustainability*; Thomas Telford Ltd.: London, UK, 2016.
2. Petley, D. Global patterns of loss of life from landslides. *Geology* **2012**, *40*, 927–930. [[CrossRef](#)]
3. Schuster, R.L.; Highland, L.M. *Socioeconomic and Environmental Impacts of Landslides in the Western Hemisphere*; U.S. Department of the Interior: Washington, DC, USA, 2001.
4. Calvello, M.; Cascini, L.; Sorbino, G. A numerical procedure for predicting rainfall-induced movements of active landslides along pre-existing slip surfaces. *Int. J. Numer. Anal. Methods Geomech.* **2008**, *32*, 327–351. [[CrossRef](#)]
5. Saito, M. Forecasting the time of occurrence of a slope failure. In *Proceedings of the 6th International Conference on Soil Mechanics and Foundation Engineering*, Montreal, QC, Canada, 8–15 September 1965.
6. Saygili, G.; Rathje, E.M. Empirical predictive models for earthquake-induced sliding displacements of slopes. *J. Geotech. Geoenviron. Eng.* **2008**, *134*, 790–803. [[CrossRef](#)]
7. Li, G.; Sun, Y.; Qi, C. Machine learning-based constitutive models for cement-grouted coal specimens under shearing. *Int. J. Min. Sci. Technol.* **2021**, *31*, 813–823. [[CrossRef](#)]

8. Wang, L.; Wu, J.; Zhang, W.; Wang, L.; Cui, W. Efficient seismic stability analysis of embankment slopes subjected to water level changes using gradient boosting algorithms. *Front. Earth Sci.* **2021**, *9*, 807317. [[CrossRef](#)]
9. Wang, Y.; Tang, H.; Wen, T.; Ma, J. A hybrid intelligent approach for constructing landslide displacement prediction intervals. *Appl. Soft Comput.* **2019**, *81*, 105506. [[CrossRef](#)]
10. Wang, K.-L.; Lin, M.-L. Initiation and displacement of landslide induced by earthquake—A study of shaking table model slope test. *Eng. Geol.* **2011**, *122*, 106–114. [[CrossRef](#)]
11. Tang, L.; Na, S. Comparison of machine learning methods for ground settlement prediction with different tunneling datasets. *J. Rock Mech. Geotech. Eng.* **2021**, *13*, 1274–1289. [[CrossRef](#)]
12. Agapiou, A. Remote sensing heritage in a petabyte-scale: Satellite data and heritage Earth Engine© applications. *Int. J. Digit. Earth* **2017**, *10*, 85–102. [[CrossRef](#)]
13. Casagli, N.; Frodella, W.; Morelli, S.; Tofani, V.; Ciampalini, A.; Intrieri, E.; Raspini, F.; Rossi, G.; Tanteri, L.; Lu, P. Spaceborne, UAV and ground-based remote sensing techniques for landslide mapping, monitoring and early warning. *Geoenviron. Disasters* **2017**, *4*, 9. [[CrossRef](#)]
14. Wang, T.; Zhang, Q.; Wu, Z. A Deep-Learning-Facilitated, Detection-First Strategy for Operationally Monitoring Localized Deformation with Large-Scale InSAR. *Remote Sens.* **2023**, *15*, 2310. [[CrossRef](#)]
15. Bekaert, D.P.; Handwerger, A.L.; Agram, P.; Kirschbaum, D.B. InSAR-based detection method for mapping and monitoring slow-moving landslides in remote regions with steep and mountainous terrain: An application to Nepal. *Remote Sens. Environ.* **2020**, *249*, 111983. [[CrossRef](#)]
16. Udin, W.S.; Norazami, N.A.S.; Sulaiman, N.; Zaudin, N.C.; Ma'ail, S.; Nor, A.M. UAV based multi-spectral imaging system for mapping landslide risk area along Jeli-Gerik highway, Jeli, Kelantan. In Proceedings of the 2019 IEEE 15th International Colloquium on Signal Processing & Its Applications (CSPA), Penang, Malaysia, 8–9 March 2019; IEEE: New York, NY, USA, 2019.
17. Tehrani, F.S.; Calvello, M.; Liu, Z.; Zhang, L.; Lacasse, S. Machine learning and landslide studies: Recent advances and applications. *Nat. Hazards* **2022**, *114*, 1197–1245. [[CrossRef](#)]
18. Eraslan, G.; Avsec, Ž.; Gagneur, J.; Theis, F.J. Deep learning: New computational modelling techniques for genomics. *Nat. Rev. Genet.* **2019**, *20*, 389–403. [[CrossRef](#)] [[PubMed](#)]
19. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [[CrossRef](#)] [[PubMed](#)]
20. Mohanty, S.P.; Czakon, J.; Kaczmarek, K.A.; Pyskir, A.; Tarasiewicz, P.; Kunwar, S.; Rohrbach, J.; Luo, D.; Prasad, M.; Flier, S.; et al. Deep learning for understanding satellite imagery: An experimental survey. *Front. Artif. Intell.* **2020**, *3*, 534696. [[CrossRef](#)] [[PubMed](#)]
21. Nguyen, T.T.; Hoang, T.D.; Pham, M.T.; Vu, T.T.; Nguyen, T.H.; Huynh, Q.T.; Jo, J. Monitoring agriculture areas with satellite images and deep learning. *Appl. Soft Comput.* **2020**, *95*, 106565. [[CrossRef](#)]
22. Schultz, M.G.; Betancourt, C.; Gong, B.; Kleinert, F.; Langguth, M.; Leufen, L.H.; Mozaffari, A.; Stadler, S. Can deep learning beat numerical weather prediction? *Philos. Trans. R. Soc. A* **2021**, *379*, 20200097. [[CrossRef](#)] [[PubMed](#)]
23. Salman, A.G.; Kanigoro, B.; Heryadi, Y. Weather forecasting using deep learning techniques. In Proceedings of the 2015 International Conference on Advanced Computer Science and Information Systems (ICACSIS), Depok, Indonesia, 10–11 October 2015; IEEE: New York, NY, USA, 2015.
24. Van Dao, D.; Jaafari, A.; Bayat, M.; Mafi-Gholami, D.; Qi, C.; Moayedi, H.; Van Phong, T.; Ly, H.B.; Le, T.T.; Trinh, P.T.; et al. A spatially explicit deep learning neural network model for the prediction of landslide susceptibility. *Catena* **2020**, *188*, 104451.
25. Azarafza, M.; Azarafza, M.; Akgün, H.; Atkinson, P.M.; Derakhshani, R. Deep learning-based landslide susceptibility mapping. *Sci. Rep.* **2021**, *11*, 24112. [[CrossRef](#)]
26. Bui, D.T.; Tsangaratos, P.; Nguyen, V.T.; Van Liem, N.; Trinh, P.T. Comparing the prediction performance of a Deep Learning Neural Network model with conventional machine learning models in landslide susceptibility assessment. *Catena* **2020**, *188*, 104426. [[CrossRef](#)]
27. Nguyen, B.-Q.-V.; Kim, Y.-T. Landslide spatial probability prediction: A comparative assessment of naive Bayes, ensemble learning, and deep learning approaches. *Bull. Eng. Geol. Environ.* **2021**, *80*, 4291–4321. [[CrossRef](#)]
28. Ghorbanzadeh, O.; Blaschke, T.; Gholamnia, K.; Meena, S.R.; Tiede, D.; Aryal, J. Evaluation of different machine learning methods and deep-learning convolutional neural networks for landslide detection. *Remote Sens.* **2019**, *11*, 196. [[CrossRef](#)]
29. Ma, W.; Liu, Z.; Kudyshev, Z.A.; Boltasseva, A.; Cai, W.; Liu, Y. Deep learning for the design of photonic structures. *Nat. Photonics* **2021**, *15*, 77–90. [[CrossRef](#)]
30. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* **2012**, *25*, 84–90. [[CrossRef](#)]
31. LeCun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE* **1998**, *86*, 2278–2324. [[CrossRef](#)]
32. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016.
33. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014.
34. Girshick, R. Fast r-cnn. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015.

35. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster r-cnn: Towards real-time object detection with region proposal networks. *Adv. Neural Inf. Process. Syst.* **2015**, *28*, 91–99. [[CrossRef](#)] [[PubMed](#)]
36. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, real-time object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016.
37. Redmon, J.; Farhadi, A. YOLO9000: Better, faster, stronger. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017.
38. Redmon, J.; Farhadi, A. Yolov3: An incremental improvement. *arXiv* **2018**, arXiv:1804.02767.
39. Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015.
40. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015: 18th International Conference, Munich, Germany, 5–9 October 2015*; Proceedings, Part III 18; Springer: Berlin/Heidelberg, Germany, 2015.
41. Elman, J.L. Finding structure in time. *Cogn. Sci.* **1990**, *14*, 179–211. [[CrossRef](#)]
42. Memory, L.S.-T. Long short-term memory. *Neural Comput.* **2010**, *9*, 1735–1780.
43. Chung, J.; Gulcehre, C.; Cho, K.; Bengio, Y. Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv* **2014**, arXiv:1412.3555.
44. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. *Adv. Neural Inf. Process. Syst.* **2017**. [[CrossRef](#)]
45. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial nets. *Adv. Neural Inf. Process. Syst.* **2014**. [[CrossRef](#)]
46. Kingma, D.P.; Welling, M. Auto-encoding variational bayes. *arXiv* **2013**, arXiv:1312.6114.
47. Ho, J.; Jain, A.; Abbeel, P. Denoising diffusion probabilistic models. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 6840–6851.
48. Tsoumakas, G.; Katakis, I. Multi-label classification: An overview. *Int. J. Data Warehous. Min. (IJDWM)* **2007**, *3*, 1–13. [[CrossRef](#)]
49. Milletari, F.; Navab, N.; Ahmadi, S.-A. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In Proceedings of the 2016 Fourth International Conference on 3D Vision (3DV), Stanford, CA, USA, 25–28 October 2016; IEEE: New York, NY, USA, 2016.
50. Wang, Z.; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612. [[CrossRef](#)]
51. Hore, A.; Ziou, D. Image quality metrics: PSNR vs. SSIM. In Proceedings of the 2010 20th International Conference on Pattern Recognition, Istanbul, Turkey, 23–26 August 2010; IEEE: New York, NY, USA, 2010.
52. Duchi, J.; Hazan, E.; Singer, Y. Adaptive subgradient methods for online learning and stochastic optimization. *J. Mach. Learn. Res.* **2011**, *12*, 2121–2159.
53. Tieleman, T.; Hinton, G. Divide the gradient by a running average of its recent magnitude. In *Neural Networks for Machine Learning*; Technical report; Coursera: Mountain View, CA, USA, 2017.
54. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.
55. Sonka, M.; Hlavac, V.; Boyle, R. *Image Processing, Analysis and Machine Vision*; Springer: Berlin/Heidelberg, Germany, 2013.
56. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
57. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015.
58. Lin, T.Y.; Dollár, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature pyramid networks for object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017.
59. Liu, S.; Qi, L.; Qin, H.; Shi, J.; Jia, J. Path aggregation network for instance segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018.
60. He, K.; Zhang, X.; Ren, S.; Sun, J. Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *37*, 1904–1916. [[CrossRef](#)] [[PubMed](#)]
61. Chen, L.C.; Papandreou, G.; Schroff, F.; Adam, H. Rethinking atrous convolution for semantic image segmentation. *arXiv* **2017**, arXiv:1706.05587.
62. Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018.
63. Woo, S.; Park, J.; Lee, J.Y.; Kweon, I.S. Cbam: Convolutional block attention module. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018.
64. Xie, S.; Girshick, R.; Dollár, P.; Tu, Z.; He, K. Aggregated residual transformations for deep neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017.
65. Tan, M.; Le, Q. Efficientnet: Rethinking model scaling for convolutional neural networks. In Proceedings of the International Conference on Machine Learning, PMLR, Long Beach, CA, USA, 9–15 June 2019.
66. Zhang, X.; Zhou, X.; Lin, M.; Sun, J. Shufflenet: An extremely efficient convolutional neural network for mobile devices. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018.

67. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv* **2020**, arXiv:2010.11929.
68. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.Y.; Berg, A.C. Ssd: Single shot multibox detector. In *Computer Vision—ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, 11–14 October 2016*; Proceedings, Part I 14; Springer: Berlin/Heidelberg, Germany, 2016.
69. Zhou, X.; Wang, D.; Krähenbühl, P. Objects as points. *arXiv* **2019**, arXiv:1904.07850.
70. Law, H.; Deng, J. Cornernet: Detecting objects as paired keypoints. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018.
71. Carion, N.; Massa, F.; Synnaeve, G.; Usunier, N.; Kirillov, A.; Zagoruyko, S. End-to-end object detection with transformers. In Proceedings of the European Conference on Computer Vision, Glasgow, UK, 23–28 August 2020; Springer: Berlin/Heidelberg, Germany, 2020.
72. Zhao, H.; Shi, J.; Qi, X.; Wang, X.; Jia, J. Pyramid scene parsing network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017.
73. Chen, L.C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. Semantic image segmentation with deep convolutional nets and fully connected crfs. *arXiv* **2014**, arXiv:1412.7062.
74. Chen, L.C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *40*, 834–848. [[CrossRef](#)] [[PubMed](#)]
75. Chen, L.C.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H. Encoder-decoder with atrous separable convolution for semantic image segmentation. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018.
76. Oktay, O.; Schlemper, J.; Folgoc, L.L.; Lee, M.; Heinrich, M.; Misawa, K.; Mori, K.; McDonagh, S.; Hammerla, N.Y.; Kainz, B.; et al. Attention u-net: Learning where to look for the pancreas. *arXiv* **2018**, arXiv:1804.03999.
77. Cheng, B.; Misra, I.; Schwing, A.G.; Kirillov, A.; Girdhar, R. Masked-attention mask transformer for universal image segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022.
78. Sarzynska-Wawer, J.; Wawer, A.; Pawlak, A.; Szymanowska, J.; Stefaniak, I.; Jarkiewicz, M.; Okruszek, L. Detecting formal thought disorder by deep contextualized word representations. *Psychiatry Res.* **2021**, *304*, 114135. [[CrossRef](#)]
79. Devlin, J.; Chang, M.W.; Lee, K.; Toutanova, K. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv* **2018**, arXiv:1810.04805.
80. Radford, A.; Narasimhan, K.; Salimans, T.; Sutskever, I. Improving language understanding by generative pre-training. *Preprint*, 2018; *Work in progress*.
81. Radford, A.; Wu, J.; Child, R.; Luan, D.; Amodei, D.; Sutskever, I. Language models are unsupervised multitask learners. *OpenAI Blog* **2019**, *1*, 9.
82. Brown, T.; Mann, B.; Ryder, N.; Subbiah, M.; Kaplan, J.D.; Dhariwal, P.; Neelakantan, A.; Shyam, P.; Sastry, G.; Askell, A.; et al. Language models are few-shot learners. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 1877–1901.
83. Achiam, J.; Adler, S.; Agarwal, S.; Ahmad, L.; Akkaya, I.; Aleman, F.L.; Almeida, D.; Altenschmidt, J.; Altman, S.; Anadkat, S.; et al. Gpt-4 technical report. *arXiv* **2023**, arXiv:2303.08774.
84. Pourghasemi, H.R.; Pradhan, B.; Gokceoglu, C.; Mohammadi, M.; Moradi, H.R. Application of weights-of-evidence and certainty factor models and their comparison in landslide susceptibility mapping at Haraz watershed, Iran. *Arab. J. Geosci.* **2013**, *6*, 2351–2365. [[CrossRef](#)]
85. Froude, M.J.; Petley, D.N. Global fatal landslide occurrence from 2004 to 2016. *Nat. Hazards Earth Syst. Sci.* **2018**, *18*, 2161–2181. [[CrossRef](#)]
86. Xu, Q.; Zhao, B.; Dai, K.; Dong, X.; Li, W.; Zhu, X.; Yang, Y.; Xiao, X.; Wang, X.; Huang, J. Remote sensing for landslide investigations: A progress report from China. *Eng. Geol.* **2023**, *321*, 107156. [[CrossRef](#)]
87. Ghorbanzadeh, O.; Xu, Y.; Ghamisi, P.; Kopp, M.; Kreil, D. Landslide4sense: Reference benchmark data and deep learning models for landslide detection. *arXiv* **2022**, arXiv:2206.00515. [[CrossRef](#)]
88. Casagli, N.; Intrieri, E.; Tofani, V.; Gigli, G.; Raspini, F. Landslide detection, monitoring and prediction with remote-sensing techniques. *Nat. Rev. Earth Environ.* **2023**, *4*, 51–64. [[CrossRef](#)]
89. Guzzetti, F.; Carrara, A.; Cardinali, M.; Reichenbach, P. Landslide hazard evaluation: A review of current techniques and their application in a multi-scale study, Central Italy. *Geomorphology* **1999**, *31*, 181–216. [[CrossRef](#)]
90. Pham, V.D.; Nguyen, Q.H.; Nguyen, H.D.; Pham, V.M.; Bui, Q.T. Convolutional neural network—Optimized moth flame algorithm for shallow landslide susceptible analysis. *IEEE Access* **2020**, *8*, 32727–32736. [[CrossRef](#)]
91. Mohan, A.; Singh, A.K.; Kumar, B.; Dwivedi, R. Review on remote sensing methods for landslide detection using machine and deep learning. *Trans. Emerg. Telecommun. Technol.* **2021**, *32*, e3998. [[CrossRef](#)]
92. Zhang, P.; Xu, C.; Ma, S.; Shao, X.; Tian, Y.; Wen, B. Automatic extraction of seismic landslides in large areas with complex environments based on deep learning: An example of the 2018 iburi earthquake, Japan. *Remote Sens.* **2020**, *12*, 3992. [[CrossRef](#)]
93. Guan, Y.; Yu, L.; Hao, S.; Li, L.; Zhang, X.; Hao, M. Slope Failure and Landslide Detection in Huangdao District of Qingdao City Based on an Improved Faster R-CNN Model. *GeoHazards* **2023**, *4*, 302–315. [[CrossRef](#)]

94. Tanatipuknon, A.; Aimmanee, P.; Watanabe, Y.; Murata, K.T.; Wakai, A.; Sato, G.; Hung, H.V.; Tungpimolrut, K.; Keeratittayanun, S.; Karnjana, J. Study on combining two faster R-CNN models for landslide detection with a classification decision tree to improve the detection performance. *J. Disaster Res.* **2021**, *16*, 588–595. [[CrossRef](#)]
95. Zhang, D.; Zhang, S.; Wang, H.; Ai, X.; Yi, N. Research on Landslide Detection in Remote Sensing Image Based on Improved Faster-RCNN. In Proceedings of the 2022 2nd International Conference on Computer, Control and Robotics (ICCCR), Shanghai, China, 18–20 March 2022; IEEE: New York, NY, USA, 2022.
96. Cai, J.; Zhang, L.; Dong, J.; Guo, J.; Wang, Y.; Liao, M. Automatic identification of active landslides over wide areas from time-series InSAR measurements using Faster RCNN. *Int. J. Appl. Earth Obs. Geoinf.* **2023**, *124*, 103516. [[CrossRef](#)]
97. Yang, D.; Mao, Y. Remote sensing landslide target detection method based on improved Faster R-CNN. *J. Appl. Remote Sens.* **2022**, *16*, 044521.
98. Lin, T.Y.; Goyal, P.; Girshick, R.; He, K.; Dollár, P. Focal loss for dense object detection. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017.
99. Fu, L.; Zhang, Q.; Wang, T.; Li, W.; Xu, Q.; Ge, D. Detecting slow-moving landslides using InSAR phase-gradient stacking and deep-learning network. *Front. Environ. Sci.* **2022**, *10*, 963322. [[CrossRef](#)]
100. Wang, T.; Liu, M.; Zhang, H.; Jiang, X.; Huang, Y.; Jiang, X. Landslide detection based on improved YOLOv5 and satellite images. In Proceedings of the 2021 4th International Conference on Pattern Recognition and Artificial Intelligence (PRAI), Yibin, China, 20–22 August 2021; IEEE: New York, NY, USA, 2021.
101. Guo, H.; Yi, B.; Yao, Q.; Gao, P.; Li, H.; Sun, J.; Zhong, C. Identification of Landslides in Mountainous Area with the Combination of SBAS-InSAR and Yolo Model. *Sensors* **2022**, *22*, 6235. [[CrossRef](#)] [[PubMed](#)]
102. Cheng, L.; Li, J.; Duan, P.; Wang, M. A small attentional YOLO model for landslide detection from satellite remote sensing images. *Landslides* **2021**, *18*, 2751–2765. [[CrossRef](#)]
103. Yang, K.; Li, W.; Yang, X.; Zhang, L. Improving Landslide Recognition on UAV Data through Transfer Learning. *Appl. Sci.* **2022**, *12*, 10121. [[CrossRef](#)]
104. Wu, L.; Liu, R.; Li, G.; Gou, J.; Lei, Y. Landslide Detection Methods Based on Deep Learning in Remote Sensing Images. In Proceedings of the 2022 29th International Conference on Geoinformatics, Beijing, China, 15–18 August 2022; IEEE: New York, NY, USA, 2022.
105. Zhang, W.; Liu, Z.; Yu, H.; Zhou, S.; Jiang, H.; Guo, Y. Comparison of landslide detection based on different deep learning algorithms. In Proceedings of the 2022 3rd International Conference on Geology, Mapping and Remote Sensing (ICGMRS), Zhoushan, China, 22–24 April 2022; IEEE: New York, NY, USA, 2022.
106. Ju, Y.; Xu, Q.; Jin, S.; Li, W.; Su, Y.; Dong, X.; Guo, Q. Loess landslide detection using object detection algorithms in northwest China. *Remote Sens.* **2022**, *14*, 1182. [[CrossRef](#)]
107. Li, H.; He, Y.; Xu, Q.; Deng, J.; Li, W.; Wei, Y. Detection and segmentation of loess landslides via satellite images: A two-phase framework. *Landslides* **2022**, *19*, 673–686. [[CrossRef](#)]
108. Bragagnolo, L.; Rezende, L.R.; Da Silva, R.V.; Grzybowski, J.M.V. Convolutional neural networks applied to semantic segmentation of landslide scars. *Catena* **2021**, *201*, 105189. [[CrossRef](#)]
109. Dong, Z.; An, S.; Zhang, J.; Yu, J.; Li, J.; Xu, D. L-unet: A landslide extraction model using multi-scale feature fusion and attention mechanism. *Remote Sens.* **2022**, *14*, 2552. [[CrossRef](#)]
110. Chen, X.; Yao, X.; Zhou, Z.; Liu, Y.; Yao, C.; Ren, K. DRs-UNet: A deep semantic segmentation network for the recognition of active landslides from InSAR imagery in the three rivers region of the Qinghai–Tibet Plateau. *Remote Sens.* **2022**, *14*, 1848. [[CrossRef](#)]
111. Huang, R.; Chen, T. Landslide recognition from multi-feature remote sensing data based on improved transformers. *Remote Sens.* **2023**, *15*, 3340. [[CrossRef](#)]
112. Yang, Z.; Xu, C.; Li, L. Landslide detection based on ResU-net with transformer and CBAM embedded: Two examples with geologically different environments. *Remote Sens.* **2022**, *14*, 2885. [[CrossRef](#)]
113. Wang, Z.; Sun, T.; Hu, K.; Zhang, Y.; Yu, X.; Li, Y. A Deep Learning Semantic Segmentation Method for Landslide Scene Based on Transformer Architecture. *Sustainability* **2022**, *14*, 16311. [[CrossRef](#)]
114. Yang, S.; Wang, Y.; Wang, P.; Mu, J.; Jiao, S.; Zhao, X.; Wang, Z.; Wang, K.; Zhu, Y. Automatic identification of landslides based on deep learning. *Appl. Sci.* **2022**, *12*, 8153. [[CrossRef](#)]
115. Du, B.; Zhao, Z.; Hu, X.; Wu, G.; Han, L.; Sun, L.; Gao, Q. Landslide susceptibility prediction based on image semantic segmentation. *Comput. Geosci.* **2021**, *155*, 104860. [[CrossRef](#)]
116. Lu, Z.; Peng, Y.; Li, W.; Yu, J.; Ge, D.; Han, L.; Xiang, W. An Iterative Classification and Semantic Segmentation Network for Old Landslide Detection Using High-Resolution Remote Sensing Images. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 4408813. [[CrossRef](#)]
117. Yu, B.; Xu, C.; Chen, F.; Wang, N.; Wang, L. HADeenNet: A hierarchical-attention multi-scale deconvolution network for landslide detection. *Int. J. Appl. Earth Obs. Geoinf.* **2022**, *111*, 102853. [[CrossRef](#)]
118. Yi, Y.; Zhang, W. A new deep-learning-based approach for earthquake-triggered landslide detection from single-temporal RapidEye satellite imagery. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *13*, 6166–6176. [[CrossRef](#)]

119. Liu, X.; Peng, Y.; Lu, Z.; Li, W.; Yu, J.; Ge, D.; Xiang, W. Feature-Fusion Segmentation Network for Landslide Detection Using High-Resolution Remote Sensing Images and Digital Elevation Model Data. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 4500314. [[CrossRef](#)]
120. Akgun, A. A comparison of landslide susceptibility maps produced by logistic regression, multi-criteria decision, and likelihood ratio methods: A case study at İzmir, Turkey. *Landslides* **2012**, *9*, 93–106. [[CrossRef](#)]
121. Chen, Y.; Wei, Y.; Wang, Q.; Chen, F.; Lu, C.; Lei, S. Mapping post-earthquake landslide susceptibility: A U-Net like approach. *Remote Sens.* **2020**, *12*, 2767. [[CrossRef](#)]
122. Jiang, Z.; Wang, M.; Liu, K. Comparisons of convolutional neural network and other machine learning methods in landslide susceptibility assessment: A case study in Pingwu. *Remote Sens.* **2023**, *15*, 798. [[CrossRef](#)]
123. Wang, D.; Yang, R.H.; Wang, X.; Li, S.D.; Tan, J.X.; Zhang, S.Q.; Wei, S.Y.; Wu, Z.Y.; Chen, C.; Yang, X.X. Evaluation of deep learning algorithms for landslide susceptibility mapping in an alpine-gorge area: A case study in Jiuzhaigou County. *J. Mt. Sci.* **2023**, *20*, 484–500. [[CrossRef](#)]
124. Bao, S.; Liu, J.; Wang, L.; Zhao, X. Application of transformer models to landslide susceptibility mapping. *Sensors* **2022**, *22*, 9104. [[CrossRef](#)] [[PubMed](#)]
125. Zhou, Y.; Hussain, M.A.; Chen, Z. Landslide susceptibility mapping with feature fusion transformer and machine learning classifiers incorporating displacement velocity along Karakoram highway. *Geocarto Int.* **2023**, *38*, 2292752. [[CrossRef](#)]
126. Mutlu, B.; Nefeslioglu, H.A.; Sezer, E.A.; Akcayol, M.A.; Gokceoglu, C. An experimental research on the use of recurrent neural networks in landslide susceptibility mapping. *ISPRS Int. J. Geo-Inf.* **2019**, *8*, 578. [[CrossRef](#)]
127. Wang, Y.; Fang, Z.; Wang, M.; Peng, L.; Hong, H. Comparative study of landslide susceptibility mapping with different recurrent neural networks. *Comput. Geosci.* **2020**, *138*, 104445. [[CrossRef](#)]
128. Yi, Y.; Zhang, W.; Xu, X.; Zhang, Z.; Wu, X. Evaluation of neural network models for landslide susceptibility assessment. *Int. J. Digit. Earth* **2022**, *15*, 934–953. [[CrossRef](#)]
129. Xiao, L.; Zhang, Y.; Peng, G. Landslide susceptibility assessment using integrated deep learning algorithm along the China-Nepal highway. *Sensors* **2018**, *18*, 4436. [[CrossRef](#)] [[PubMed](#)]
130. Ji, J.; Zhou, Y.; Cheng, Q.; Jiang, S.; Liu, S. Landslide Susceptibility Mapping Based on Deep Learning Algorithms Using Information Value Analysis Optimization. *Land* **2023**, *12*, 1125. [[CrossRef](#)]
131. Li, W.; Fang, Z.; Wang, Y. Stacking ensemble of deep learning methods for landslide susceptibility mapping in the Three Gorges Reservoir area, China. *Stoch. Environ. Res. Risk Assess.* **2021**, *36*, 2207–2228. [[CrossRef](#)]
132. Chen, C.; Fan, L. CNN-LSTM-attention deep learning model for mapping landslide susceptibility in Kerala, India. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2022**, *10*, 25–30. [[CrossRef](#)]
133. Yuan, R.; Chen, J. A hybrid deep learning method for landslide susceptibility analysis with the application of InSAR data. *Nat. Hazards* **2022**, *114*, 1393–1426. [[CrossRef](#)]
134. Yuan, R.; Chen, J. A novel method based on deep learning model for national-scale landslide hazard assessment. *Landslides* **2023**, *20*, 2379–2403. [[CrossRef](#)]
135. Wang, H.; Zhang, L.; Luo, H.; He, J.; Cheung, R.W.M. AI-powered landslide susceptibility assessment in Hong Kong. *Eng. Geol.* **2021**, *288*, 106103. [[CrossRef](#)]
136. Du, J.; Yin, K.; Lacasse, S. Displacement prediction in colluvial landslides, three Gorges reservoir, China. *Landslides* **2013**, *10*, 203–218. [[CrossRef](#)]
137. Chen, H.; Zeng, Z.; Tang, H. Landslide deformation prediction based on recurrent neural network. *Neural Process. Lett.* **2015**, *41*, 169–178. [[CrossRef](#)]
138. Zhang, W.; Li, H.; Tang, L.; Gu, X.; Wang, L.; Wang, L. Displacement prediction of Jiuxianping landslide using gated recurrent unit (GRU) networks. *Acta Geotech.* **2022**, *17*, 1367–1382. [[CrossRef](#)]
139. Xi, N.; Yang, Q.; Sun, Y.; Mei, G. Machine Learning Approaches for Slope Deformation Prediction Based on Monitored Time-Series Displacement Data: A Comparative Investigation. *Appl. Sci.* **2023**, *13*, 4677. [[CrossRef](#)]
140. Yang, B.; Yin, K.; Lacasse, S.; Liu, Z. Time series analysis and long short-term memory neural network to predict landslide displacement. *Landslides* **2019**, *16*, 677–694. [[CrossRef](#)]
141. Zhang, Y.G.; Tang, J.; He, Z.Y.; Tan, J.; Li, C. A novel displacement prediction method using gated recurrent unit model with time series analysis in the Erdaohe landslide. *Nat. Hazards* **2021**, *105*, 783–813. [[CrossRef](#)]
142. Xie, P.; Zhou, A.; Chai, B. The application of long short-term memory (LSTM) method on displacement prediction of multifactor-induced landslides. *IEEE Access* **2019**, *7*, 54305–54311. [[CrossRef](#)]
143. Lin, Z.; Ji, Y.; Sun, X. Landslide Displacement Prediction Based on CEEMDAN Method and CNN-BiLSTM Model. *Sustainability* **2023**, *15*, 10071. [[CrossRef](#)]
144. Nava, L.; Carraro, E.; Reyes-Carmona, C.; Puliero, S.; Bhuyan, K.; Rosi, A.; Monserrat, O.; Floris, M.; Meena, S.R.; Galve, J.P.; et al. Landslide displacement forecasting using deep learning and monitoring data across selected sites. *Landslides* **2023**, *20*, 2111–2129. [[CrossRef](#)]
145. Li, L.M.; Wang, C.Y.; Wen, Z.Z.; Gao, J.; Xia, M.F. Landslide displacement prediction based on the ICEEMDAN, ApEn and the CNN-LSTM models. *J. Mt. Sci.* **2023**, *20*, 1220–1231. [[CrossRef](#)]
146. Jiang, Y.; Luo, H.; Xu, Q.; Lu, Z.; Liao, L.; Li, H.; Hao, L. A graph convolutional incorporating GRU network for landslide displacement forecasting based on spatiotemporal analysis of GNSS observations. *Remote Sens.* **2022**, *14*, 1016. [[CrossRef](#)]

147. Kuang, P.; Li, R.; Huang, Y.; Wu, J.; Luo, X.; Zhou, F. Landslide Displacement Prediction via Attentive Graph Neural Network. *Remote Sens.* **2022**, *14*, 1919. [[CrossRef](#)]
148. Yang, B.; Guo, Z.; Wang, L.; He, J.; Xia, B.; Vakily, S. Updated Global Navigation Satellite System Observations and Attention-Based Convolutional Neural Network–Long Short-Term Memory Network Deep Learning Algorithms to Predict Landslide Spatiotemporal Displacement. *Remote Sens.* **2023**, *15*, 4971. [[CrossRef](#)]
149. Al-Najjar, H.A.; Pradhan, B.; Sarkar, R.; Beydoun, G.; Alamri, A. A new integrated approach for landslide data balancing and spatial prediction based on generative adversarial networks (GAN). *Remote Sens.* **2021**, *13*, 4011. [[CrossRef](#)]
150. Lee, J.H.; Sameen, M.I.; Pradhan, B.; Park, H.J. Modeling landslide susceptibility in data-scarce environments using optimized data mining and statistical methods. *Geomorphology* **2018**, *303*, 284–298. [[CrossRef](#)]
151. Reichenbach, P.; Rossi, M.; Malamud, B.D.; Mihir, M.; Guzzetti, F. A review of statistically-based landslide susceptibility models. *Earth-Sci. Rev.* **2018**, *180*, 60–91. [[CrossRef](#)]
152. Zhao, F.; Meng, X.; Zhang, Y.; Chen, G.; Su, X.; Yue, D. Landslide susceptibility mapping of karakorum highway combined with the application of SBAS-InSAR technology. *Sensors* **2019**, *19*, 2685. [[CrossRef](#)]
153. Dai, F.; Lee, C.F.; Ngai, Y.Y. Landslide risk assessment and management: An overview. *Eng. Geol.* **2002**, *64*, 65–87. [[CrossRef](#)]
154. Huang, F.; Tao, S.; Chang, Z.; Huang, J.; Fan, X.; Jiang, S.H.; Li, W. Efficient and automatic extraction of slope units based on multi-scale segmentation method for landslide assessments. *Landslides* **2021**, *18*, 3715–3731. [[CrossRef](#)]
155. Liu, Y.; Xu, C.; Huang, B.; Ren, X.; Liu, C.; Hu, B.; Chen, Z. Landslide displacement prediction based on multi-source data fusion and sensitivity states. *Eng. Geol.* **2020**, *271*, 105608. [[CrossRef](#)]
156. Tavakkoli Piralilou, S.; Shahabi, H.; Jarihani, B.; Ghorbanzadeh, O.; Blaschke, T.; Gholamnia, K.; Meena, S.R.; Aryal, J. Landslide detection using multi-scale image segmentation and different machine learning models in the higher himalayas. *Remote Sens.* **2019**, *11*, 2575. [[CrossRef](#)]
157. Kadavi, P.R.; Lee, C.-W.; Lee, S. Application of ensemble-based machine learning models to landslide susceptibility mapping. *Remote Sens.* **2018**, *10*, 1252. [[CrossRef](#)]
158. Liu, C.; Li, W.; Wu, H.; Lu, P.; Sang, K.; Sun, W.; Chen, W.; Hong, Y.; Li, R. Susceptibility evaluation and mapping of China's landslides based on multi-source data. *Nat. Hazards* **2013**, *69*, 1477–1495. [[CrossRef](#)]
159. Zhang, D.; Sun, K.; Zhang, S. An Approach to Data Modeling via Temporal and Spatial Alignment. *Processes* **2023**, *12*, 62. [[CrossRef](#)]
160. Ferretti, A.; Massonnet, D.; Monti-Guarnieri, A.; Prati, C.; Rocca, F. *Guidelines for Sar Interferometry Processing and Interpretation*; ESA Publications Division: Noordwijk, The Netherlands, 2007.
161. Lipton, Z.C. The mythos of model interpretability: In machine learning, the concept of interpretability is both important and slippery. *Queue* **2018**, *16*, 31–57. [[CrossRef](#)]
162. Zhang, Q.-S.; Zhu, S.-C. Visual interpretability for deep learning: A survey. *Front. Inf. Technol. Electron. Eng.* **2018**, *19*, 27–39. [[CrossRef](#)]
163. Lundberg, S.M.; Lee, S.-I. A unified approach to interpreting model predictions. *Adv. Neural Inf. Process. Syst.* **2017**. [[CrossRef](#)]
164. Ribeiro, M.T.; Singh, S.; Guestrin, C. “Why should i trust you?” Explaining the predictions of any classifier. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016.
165. Zhou, B.; Khosla, A.; Lapedriza, A.; Oliva, A.; Torralba, A. Learning deep features for discriminative localization. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016.
166. Selvaraju, R.R.; Cogswell, M.; Das, A.; Vedantam, R.; Parikh, D.; Batra, D. Grad-cam: Visual explanations from deep networks via gradient-based localization. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017.
167. Zhang, J.; Ma, X.; Zhang, J.; Sun, D.; Zhou, X.; Mi, C.; Wen, H. Insights into geospatial heterogeneity of landslide susceptibility based on the SHAP-XGBoost model. *J. Environ. Manag.* **2023**, *332*, 117357. [[CrossRef](#)]
168. Hacıfendioglu, K.; Demir, G.; Başağa, H.B. Landslide detection using visualization techniques for deep convolutional neural network models. *Nat. Hazards* **2021**, *109*, 329–350. [[CrossRef](#)]
169. Zhou, Y.; Wang, H.; Yang, R.; Yao, G.; Xu, Q.; Zhang, X. A novel weakly supervised remote sensing landslide semantic segmentation method: Combining cam and cyclegan algorithms. *Remote Sens.* **2022**, *14*, 3650. [[CrossRef](#)]
170. Chefer, H.; Gur, S.; Wolf, L. Transformer interpretability beyond attention visualization. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021.
171. Aslam, B.; Zafar, A.; Khalil, U. Development of integrated deep learning and machine learning algorithm for the assessment of landslide hazard potential. *Soft Comput.* **2021**, *25*, 13493–13512. [[CrossRef](#)]
172. Raissi, M.; Perdikaris, P.; Karniadakis, G.E. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *J. Comput. Phys.* **2019**, *378*, 686–707. [[CrossRef](#)]
173. Kalantar, B.; Ueda, N.; Saeidi, V.; Ahmadi, K.; Halin, A.A.; Shabani, F. Landslide susceptibility mapping: Machine and ensemble learning based on remote sensing big data. *Remote Sens.* **2020**, *12*, 1737. [[CrossRef](#)]
174. Chen, T.; Kornblith, S.; Norouzi, M.; Hinton, G. A simple framework for contrastive learning of visual representations. In Proceedings of the International Conference on Machine Learning, PMLR, Virtual Event, 13–18 July 2020.

175. Sameen, M.I.; Pradhan, B. Landslide detection using residual networks and the fusion of spectral and topographic information. *IEEE Access* **2019**, *7*, 114363–114373. [[CrossRef](#)]
176. Ou, X.; Liu, L.; Tan, S.; Zhang, G.; Li, W.; Tu, B. A hyperspectral image change detection framework with self-supervised contrastive learning pretrained model. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2022**, *15*, 7724–7740. [[CrossRef](#)]

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