MU-Net: Multi-Scale for Enhanced Landslide Susceptibility Mapping

Nguyen Gia Huy
Institute for Artificial Intelligence
VNU University of Engineering and Technology
Ha Noi, Vietnam
23020377@vnu.edu.vn

Abstract

Landslides are a significant natural hazard that causes substantial loss of life and economic damage globally. Accurate Landslide Susceptibility Mapping (LSM) is therefore essential for effective risk assessment, land-use planning, and disaster mitigation. This paper introduces a novel deep learning architecture featuring a multi-scale feature extraction module for improved landslide susceptibility mapping. Our proposed model processes input data through four parallel convolutional pathways: a 1x1 convolution to capture fine-grained pixel-level information, a standard 3x3 convolution for local features, a 3x3 dilated convolution to capture wider contextual information without losing resolution, and a max-pooling-based branch to extract the most salient features. These parallel feature maps are then concatenated, creating a rich, hierarchical representation that captures information across multiple receptive fields simultaneously. The model was trained and validated using the Landslide4Sense benchmark, from which we derived 6 key conditioning factors including spectral bands, NDVI, and topographic data. Experimental results demonstrate that our multi-scale approach outperforms baseline deep learning methods, achieving an F1-score of 0.6075. The proposed framework offers a more robust and reliable tool for identifying landslide-prone areas, contributing to more effective risk management strategies.

1 Introduction

Landslides are among the most destructive and widespread geological hazards, posing significant threats to human life, infrastructure, and the environment. Every year, these events are responsible for thousands of fatalities and billions of dollars in economic damages worldwide. Effective disaster risk management and land-use planning depend critically on the ability to accurately identify and map areas susceptible to landslides. Landslide Susceptibility Mapping (LSM) is the process of estimating the spatial likelihood of landslide occurrences based on an analysis of a region's geo-environmental conditioning factors, such as slope, geology, land cover, and proximity to faults. The resulting susceptibility maps are indispensable tools for policymakers, engineers, and emergency responders to implement targeted mitigation strategies and safeguard vulnerable communities.

The societal and economic imperatives for accurate LSM cannot be overstated. High-fidelity susceptibility maps enable the development of early warning systems, inform the creation of zoning regulations that restrict construction in high-risk areas, and guide the allocation of resources for reinforcing critical infrastructure like roads and pipelines. Furthermore, in the context of climate change, which is predicted to increase the frequency and intensity of landslide-triggering events like extreme rainfall, the need for precise and reliable predictive models is more urgent than ever. An improvement in mapping accuracy, even by a few percentage points, can translate into significantly better-protected communities and more resilient infrastructure.

Despite its importance, producing accurate LSM is a complex and challenging task due to several inherent difficulties:

- **Complex Non-linear Relationships:** The intricate, non-linear interactions between multiple landslide conditioning factors pose a significant challenge for traditional statistical models.
- **Scale-Dependent Feature Representation:** The influence of conditioning factors is scale-dependent, with some being critical locally and others regionally, a context often missed by single-scale analysis.
- **Data Imbalance and Quality:** Ground-truth landslide inventory data is often incomplete, spatially biased, and suffers from severe class imbalance, with landslide pixels being vastly outnumbered by non-landslide pixels.

To address the aforementioned challenges in Landslide Susceptibility Mapping (LSM), this research makes two primary contributions that lead to a more accurate and robust predictive model:

- **Integration of a Multi-Scale Feature Extraction Module:** We propose a novel architecture that incorporates a specialized module for multi-scale feature analysis. This module processes input data through parallel convolutional pathways with varying receptive fields, utilizing standard, dilated, and 1x1 convolutions. This approach directly confronts the problem of scale-dependent feature representation by allowing the model to simultaneously learn fine-grained local details and broader regional context . By fusing these hierarchical features, our model develops a more comprehensive understanding of the factors contributing to landslides, leading to a significant improvement in predictive performance over standard single-scale models.
- **Strategic Implementation of a Hybrid Loss Function:** To tackle the severe class imbalance inherent in landslide data, where non-landslide pixels vastly outnumber landslide pixels, we employ a hybrid loss function that combines Binary Cross-Entropy (BCE) and Dice Loss. While BCE ensures pixel-level accuracy across the entire map, the Dice Loss component specifically focuses on maximizing the spatial overlap of the rare landslide class. This prevents the model from converging on a trivial solution and significantly improves the structural integrity and accuracy of the predicted landslide-prone areas.

2 Related Work

2.1 Efficient Network Architecture

The U-Net architecture is a foundational model for semantic segmentation, featuring a symmetric encoder-decoder structure. The encoder path captures high-level semantic context through a series of convolutions and downsampling operations, while the decoder path reconstructs a full-resolution map. U-Net's key innovation is its use of long-range skip connections, which fuse high-resolution spatial details from the encoder with the abstract feature maps in the decoder, enabling precise localization. To enhance the training of deeper and more powerful segmentation networks, the ResU-Net architecture was proposed. It improves upon the standard U-Net by replacing its plain convolutional blocks with the residual blocks pioneered by ResNet. These blocks introduce identity shortcut connections that facilitate better gradient flow and allow the network to learn more complex feature representations without suffering from the degradation problem common in very deep models. This modification effectively combines the structural advantages of U-Net with the superior training dynamics of residual learning.

2.2 Encoder-decoder structure

Our network adopts the symmetric encoder-decoder paradigm, a foundational structure in semantic segmentation pioneered by models like U-Net. The architecture consists of two main pathways. The first is an encoder (the downsampling path), which progressively reduces the spatial dimensions of the input to capture abstract, semantic information—essentially determining what is in the image. The second is a decoder (the upsampling path), which systematically restores the spatial resolution to determine where these features are located. To ensure precise localization and boundary delineation, this architecture is enhanced with skip connections that feed high-resolution feature maps from the encoder directly to the decoder. This mechanism effectively combines the "what" information from deep layers with the "where" information from shallow layers, a synergy that is crucial for high-quality segmentation. We selected this architecture as a robust and effective baseline for our work.

3 Method

3.1 Dataset and Preprocessing

The Landslide4Sense dataset is designed as a multi-source benchmark for training deep learning (DL) models in landslide detection. Given the challenges posed by small or homogeneous datasets, this benchmark incorporates data from four diverse geographic regions, including the Iburi-Tobu Area of Hokkaido, the Kodagu District of Karnataka, the Rasuwa District of Bagmati, and Western Taitung County—located in Japan, India, Nepal, and Taiwan, China, respectively. Among these, the landslides in Japan and Nepal are of the earthquake-induced type (the April 2015 Nepal earthquake and the August 2018 Hokkaido earthquake), while those in India and Taiwan, China, are triggered by heavy rainfall and Typhoon. The Landslide4Sense dataset is divided into three splits:

- **Training set:** 3799 images
- **Validation set:** 245 images
- **Test set:** 800 images

Each image patch is composed of 14 spectral and auxiliary bands, as detailed below:

Band	Description
B1	Coastal aerosol
B2	Blue
B3	Green
B4	Red
B5	Vegetation Red Edge 1
B6	Vegetation Red Edge 2
B7	Vegetation Red Edge 3
B8	Near Infrared (NIR)
B9	Narrow Near Infrared
B10	Shortwave Infrared - Cirrus
B11	Shortwave Infrared 1
B12	Shortwave Infrared 2
B13	Slope (from ALOS PALSAR)
B14	Digital Elevation Model (DEM from ALOS PALSAR)

Additional details: All bands are resampled to approximately 10 meters per pixel resolution. Each image patch has a spatial size of 128×128 pixels with pixel-wise labels.

The input data for the model consisted of a multi-channel raster stack composed of six selected conditioning factors: Red, Green, Blue, NDVI, Slope, and Elevation. These features were chosen based on their high influence and low inter-correlation, as determined through a feature importance analysis conducted during the preprocessing phase. NDVI was computed from the Sentinel-2 bands using the formula NDVI $= \frac{B8-B4}{B8+B4}$, where $B8$ is the near-infrared (NIR) band and $B4$ is the red band. All raster layers were co-registered and resampled to a uniform spatial resolution to ensure perfect pixel-level alignment.

To handle the diverse scales and units across the different factors, each channel was independently standardized using Z-score normalization:

$$X_{\text{std}} = \frac{X - \mu}{\sigma}$$

3.2 Model Architecture

A primary challenge in landslide segmentation is the significant variation in object scale. Large, complex landslides require a broad contextual view to be fully understood, while small landslides—which are particularly difficult to detect—can be lost or misclassified without a focus on fine-grained local details. Standard U-Net architectures, while effective, pass features through their skip connections directly, providing the decoder with raw, single-scale information from the encoder.

To address this limitation, we introduce a novel component which we term the *Multi-scale Connection*. Instead of a direct skip connection, our module is strategically placed on the pathway between the encoder and the decoder. It intercepts the feature maps coming from the encoder before they are fused with the decoder's upsampled features. The purpose of this module is to enrich the skip-connection data, transforming it from a simple, low-level feature map into a rich, multi-scale representation.

Multi-Scale Connection Module

Our proposed **Multi-Scale Connection (MSC)** module is designed to enhance the feature representation within the U-Net's skip pathways. Instead of a direct connection, the MSC module intercepts features from the encoder and enriches them with information captured at multiple receptive fields before they are fused with the decoder's feature maps.

Mechanism As illustrated in the Upsample block of our implementation, for a given skip connection feature map x_{enc} with C channels, the MSC module processes it through four parallel streams. To maintain computational efficiency, each stream is designed to output a feature map with $C/4$ channels. The streams are:

- **Pixel-level Stream (S_0):** A 1×1 convolution captures fine-grained features and performs channel-wise feature recalibration.
- **Local Context Stream (S_1):** A standard 3×3 convolution extracts local textures and patterns, analogous to a typical convolutional layer.
- **Wide Context Stream (S_2):** A 3×3 dilated convolution with a dilation rate of 3 significantly expands the receptive field without downsampling. This allows the model to incorporate broader contextual information, which is crucial for identifying the full extent of large landslides.
- **Salient Feature Stream (S_3):** A 3×3 max-pooling layer (with stride = 1 and padding = 1 to preserve spatial dimensions) identifies the most dominant activations in a local region, which are then processed by a 1×1 convolution. This branch focuses on the most salient signals.

The outputs of the four streams, each with $C/4$ channels, are concatenated along the channel dimension. This results in a final, enriched feature map x_{msc} with the original channel count C, but now densely packed with a composite of fine details, local context, broad context, and salient features.

Formally, let $x_{\text{enc}}^{(i)}$ be the feature map from the i-th encoder's skip connection, and let $x_{\text{dec}}^{(i+1)}$ be the output from the deeper $(i+1)$-th decoder block. The entire operation within the Upsample block at decoder stage i, which produces the final output $x_{\text{dec}}^{(i)}$, can be described as a sequential process:

Upsample Deeper Features: First, the features from the deeper decoder layer, $x_{\text{dec}}^{(i+1)}$, are upsampled by a factor of 2 using bilinear interpolation to match the spatial dimensions of the skip connection. Let's denote the result as $x_{\text{up}}^{(i)}$:

$$x_{\text{up}}^{(i)} = \text{Interpolate}(x_{\text{dec}}^{(i+1)}) \tag{1}$$

Enrich Skip Connection: Concurrently, the corresponding high-resolution features from the encoder, $x_{\text{enc}}^{(i)}$, are processed by the Multi-Scale Connection (MSC) module to create an enriched, multi-scale representation:

$$x_{\text{msc}}^{(i)} = \text{MSC}(x_{\text{enc}}^{(i)}) = \text{Concat}\left(S_0(x_{\text{enc}}^{(i)}),\ S_1(x_{\text{enc}}^{(i)}),\ S_2(x_{\text{enc}}^{(i)}),\ S_3(x_{\text{enc}}^{(i)})\right) \tag{2}$$

Fuse and Refine: The upsampled features $x_{\text{up}}^{(i)}$ and the enriched skip features $x_{\text{msc}}^{(i)}$ are then concatenated. This fused tensor is immediately passed through a final Residual Convolutional (ResConv) block, which serves to intelligently integrate the semantic information from the decoder path with the detailed, multi-scale information from the enhanced skip connection. The output of this ResConv block is the final output of the decoder stage, $x_{\text{dec}}^{(i)}$.

Final Expression: This entire sequence is captured in a single, nested operation that defines the output of the i-th decoder stage:

$$x_{\text{dec}}^{(i)} = \text{ResConv}\left(\text{Concat}\left(\text{Interpolate}(x_{\text{dec}}^{(i+1)}),\ \text{MSC}(x_{\text{enc}}^{(i)})\right)\right) \tag{3}$$

This enhanced information flow ensures that at each stage of reconstruction, the features are not just naively combined but are actively refined by a dedicated residual block. This leads to a more robust fusion of information and contributes to improved segmentation accuracy, especially for objects with significant scale variation like landslides.

3.3 Formulation of loss function

A major challenge in data-driven Landslide Susceptibility Mapping (LSM) is severe class imbalance—landslide pixels typically comprise less than 1–2% of any study area. This imbalance can significantly hinder model performance if not properly addressed.

Pitfalls of Standard Loss Functions: Training with a standard Binary Cross-Entropy (BCE) loss often leads to the *accuracy paradox*, where models predict only the dominant "non-landslide" class and still achieve high accuracy—yet fail to identify actual landslides. Furthermore, the gradient contributions from the majority class dominate the learning process, causing the model to largely ignore rare but critical landslide pixels.

The Binary Cross-Entropy (BCE) loss is defined as:

$$\mathcal{L}_{\text{BCE}} = -\frac{1}{N}\sum_{i=1}^{N}\left[y_i \log(\hat{y}_i) + (1 - y_i)\log(1 - \hat{y}_i)\right]$$

where:

- N is the number of samples (or pixels in segmentation tasks),
- $y_i \in \{0, 1\}$ is the ground truth label for the i-th sample,
- $\hat{y}_i \in (0, 1)$ is the predicted probability for the i-th sample.

Dice Loss as a Solution: Dice Loss mitigates these issues by directly measuring the spatial overlap between predicted and ground-truth landslide regions. Unlike BCE, it focuses on the minority (landslide) class, and strongly penalizes mislocalization (e.g., predicting "no landslide" everywhere results in a Dice score of 0 and maximum loss).

The Dice Loss is derived from the Dice Similarity Coefficient (DSC) and is defined as:

$$\mathcal{L}_{\text{Dice}} = 1 - \frac{2\sum_{i=1}^{N} y_i \hat{y}_i + \epsilon}{\sum_{i=1}^{N} y_i + \sum_{i=1}^{N} \hat{y}_i + \epsilon}$$

where:

- N is the number of samples (e.g., pixels in an image),
- $y_i \in \{0, 1\}$ is the ground truth label for the i-th sample,
- $\hat{y}_i \in [0, 1]$ is the predicted probability for the i-th sample,
- ϵ is a small constant added for numerical stability.

Combined BCE + Dice Loss: We adopt a combined BCE-Dice loss to balance pixel-level accuracy (via BCE) with robustness to class imbalance (via Dice). This synergy helps the model learn meaningful patterns in landslide-prone areas while avoiding trivial solutions, yielding predictions that are both statistically accurate and practically reliable for risk assessment.

$$\mathcal{L}_{\text{Hybrid}} = \mathcal{L}_{\text{BCE}} + \mathcal{L}_{\text{Dice}}$$

4 Experiments

4.1 Training Strategy

We compared four architectures (U-Net, ResU-Net, and our proposed MU-Net, MResU-Net) trained from scratch on 3,799 image patches. All models were optimized using the Adam optimizer with an initial learning rate of 1×10^{-3}, and a combined Binary Cross-Entropy (BCE) and Dice loss function. A `ReduceLROnPlateau` scheduler was employed to halve the learning rate after 2 epochs of stagnant validation loss, and an early stopping mechanism with a patience of 5 epochs was used to terminate training, with a maximum of 30 epochs. All experiments were conducted on a PC equipped with an Intel i5-13500HX CPU, an NVIDIA RTX 4050 Laptop GPU, and 16 GB of RAM.

4.2 Evaluation Metrics

To quantitatively evaluate the performance of my method and related works, this article selects six commonly used evaluation metrics based on the binary confusion matrix, namely, overall accuracy, precision, recall, F1-score, Kappa coefficient, and mean intersection over union (MIoU) index.

TP, FP, TN, and FN denote the number of true positives, false positives, true negatives, and false negatives, respectively. The variable n represents the total number of samples.

Overall accuracy is a commonly used evaluation index and is generally defined as the proportion of the number of samples with correct classification to the total number of samples.

$$\text{OA} = \frac{TP + TN}{TP + TN + FP + FN}$$

However, the index is not applicable in datasets with unbalanced samples. In binary classification problems such as landslide recognition, precision and recall are usually the most commonly used performance evaluation indicators. The **Precision** metric calculates how specific models are in landslide detection, and **Recall** represents how many landslide pixels are correctly detected. The **F1** measure is a combined measure between Precision and Recall. The formulas are as follows:

$$\text{Precision } = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{4}$$

$$\text{Recall } = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{5}$$

$$F1 - \text{Score } = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}. \tag{6}$$

The **Kappa coefficient** is an index used to measure classification accuracy, which quantitatively evaluates the consistency between classification results and the ground truth labels. Its value typically lies between 0 and 1, and is generally interpreted as follows:

- 0.0–0.20: Slight agreement
- 0.21–0.40: Fair agreement
- 0.41–0.60: Moderate agreement
- 0.61–0.80: Substantial agreement
- 0.81–1.00: Almost perfect agreement

Therefore, a Kappa value greater than 0.8 is considered to indicate strong consistency. The Kappa coefficient is defined as:

$$\kappa = \frac{p_0 - p_e}{1 - p_e} \tag{7}$$

where p_0 refers to the overall accuracy, and p_e is the expected accuracy by chance, computed as:

$$p_e = \frac{(\mathrm{TP}+\mathrm{FN})(\mathrm{TP}+\mathrm{FP}) + (\mathrm{FP}+\mathrm{TN})(\mathrm{FN}+\mathrm{TN})}{n^2} \tag{8}$$

MIoU is the standard measure of semantic segmentation. It starts by finding the intersection ratio of each category on the two sets and then averaging the sets. In semantic segmentation, these two sets are the ground truth data and the predicted segmentation data. Ideally, the true value and the predicted value are exactly the same, i.e., the MIoU value is 1. In this study, the formula is as follows:

$$\mathrm{MIoU} = \frac{1}{C}\sum_{i=1}^{C} \frac{TP_i}{TP_i + FP_i + FN_i}$$

4.3 Testing results

The performance of our proposed MU-Net, ResMU-Net, was compared against two ablation models: a standard U-Net and a U-Net with residual blocks (ResU-Net). All models were trained and evaluated under identical conditions. The results on the test set are summarized in Table 1.

Table 1: Comparison of Segmentation Models on Evaluation Metrics

Model	Overall Accuracy	Precision	Recall	F1-Score	Kappa	Mean IoU
U-Net	0.9807	0.4913	0.6809	0.5707	0.5611	0.3993
MU-Net	0.9779	0.4463	**0.7051**	0.5466	0.5359	0.3761
ResU-Net	0.9837	0.5579	0.6589	0.6042	0.5960	0.4329
ResMU-Net	**0.9847**	**0.5884**	0.6278	**0.6075**	**0.5997**	**0.4362**

The quantitative results demonstrate that our proposed ResMU-Net achieves the best overall performance, leading in four of the six metrics, including the most critical F1-Score and Mean IoU.

A deeper analysis of the results reveals a clear and compelling story about the synergy between residual connections and our proposed Multiscale module.

1. **The Accuracy Paradox and the Importance of IoU/F1-Score**
   First, we note that all models achieve very high accuracy (>97.7%). This is a classic symptom of the class imbalance problem in landslide mapping, where a model can appear accurate by simply predicting the majority "non-landslide" class. This highlights the limited utility of accuracy as a metric and underscores the importance of the F1-Score and IoU, which better reflect the model's ability to correctly identify the rare landslide class.
2. **The Ambiguous Impact of the Multiscale Module on a Weak Backbone (U-Net vs. MU-Net)**
   Interestingly, simply adding our Multiscale module to the standard U-Net (creating MU-Net) does not improve performance; in fact, it slightly degrades it (F1-Score drops from 0.57 to 0.54). The MU-Net does show the highest Recall, suggesting it is more aggressive in identifying potential landslides, but this comes at a steep cost to Precision (dropping from 0.49 to 0.44), indicating a high number of false positives.

   This crucial result suggests that the Multiscale module adds complexity that cannot be properly leveraged by a weak feature extractor. The standard U-Net backbone may not provide features that are semantically rich enough for the multi-scale analysis to be beneficial; instead, the added complexity might introduce noise or lead the model to overfit on simple, low-level textures, hence the higher recall but lower precision and overall score.

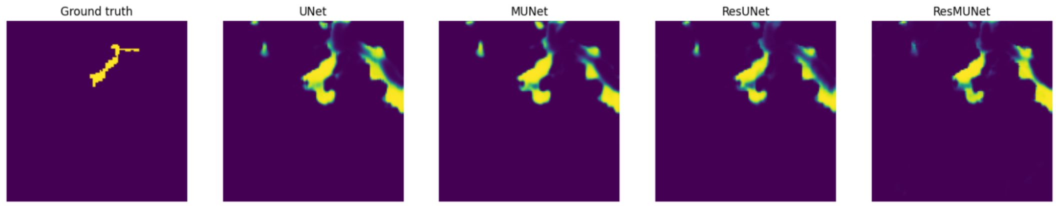

Figure 1: Visual comparison of landslide segmentation results on a sample from the test set

3. **The Synergy of Residual Blocks and the Multiscale Module (ResU-Net vs. ResMU-Net)** The true value of our contribution is revealed when comparing ResU-Net with our final ResMU-Net. With the strong, semantically rich features provided by the residual backbone, the Multiscale module is now able to perform its function effectively.

   ResMU-Net achieves the highest Precision, F1-Score, and IoU. This demonstrates that when the Multiscale module receives high-quality features on the skip connection, it can successfully enrich them. The fine-grained streams (1x1, 3x3 conv) help to refine boundaries, while the broad-context streams (dilated conv, pooling) help to differentiate true landslides from similarly textured terrain, thus boosting precision.

   The improvement is not a massive leap over ResU-Net, but it is a consistent and clear enhancement on the most challenging metrics. It represents the final, crucial refinement that pushes the model to peak performance.

Analysis of Visual Results The figure shows the ground truth alongside predictions from U-Net, MU-Net, ResU-Net, and our proposed ResMU-Net.

- **U-Net:** Produces a smeared, fragmented prediction with poor boundary alignment and multiple false positives.
- **MU-Net:** Adds multiscale complexity to a weak backbone, leading to over-segmentation and disconnected blobs.
- **ResU-Net:** Significantly improves spatial coherence and reduces noise, but the boundaries are still overly smooth.
- **ResMU-Net:** Best captures the fine structure of the landslide with sharp boundaries and minimal noise, confirming the synergy between residual and multiscale modules.

5 Conclusion

In this paper, we introduced a novel deep learning architecture designed to enhance the accuracy of Landslide Susceptibility Mapping. We addressed two critical challenges in this domain: the complex, scale-dependent nature of landslide conditioning factors and the severe class imbalance inherent in landslide datasets. Our primary contribution is the integration of a Multiscale Connection module within a residual U-Net framework. This module processes features from the encoder's skip connections through four parallel streams—using 1x1, standard, and dilated convolutions, alongside a max-pooling path—to capture a rich hierarchy of spatial information, from fine-grained details to broad contextual patterns. Furthermore, we employed a hybrid loss function combining Binary Cross-Entropy and Dice Loss to effectively mitigate the class imbalance problem, ensuring the model focuses on the accurate segmentation of the rare landslide class.

Our experimental results, conducted on the challenging Landslide4Sense benchmark dataset, validate the effectiveness of our approach. The ResMU-Net model consistently outperformed standard U-Net, MU-Net and ResU-Net architectures, achieving the highest scores on key segmentation metrics, including an F1-Score of 0.6075 and a Mean IoU of 0.4362. A key insight from our ablation study is the critical synergy between the residual backbone and the Multiscale Connection module; the high-quality, semantically rich features produced by the residual blocks are essential for the multiscale analysis to be effective, leading to more precise and spatially coherent predictions.

The proposed architecture provide a more robust and reliable tool for identifying landslide-prone areas, which can significantly contribute to improved risk assessment, land-use planning, and disaster mitigation strategies. Future work could explore the integration of attention mechanisms within the multiscale module to further refine feature fusion. Additionally, extending the model to incorporate temporal data could enable a transition from susceptibility mapping to near-real-time landslide detection. Finally, testing the generalizability of the model on diverse geographic regions and landslide types remains an important direction for future research.

References

[1] Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015* (pp. 234-241). Springer.

[2] Ghorbanzadeh, O., Ghamisi, P., Ghaffarian, S., & Momeni, M. (2022). Landslide4Sense: A Multi-Source Benchmark Dataset for Landslide Detection Using Deep Learning. *IEEE Transactions on Geoscience and Remote Sensing*, **60**, 1-17.

[3] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 770-778).

[4] Zhang, Z., Liu, Q., & Wang, Y. (2018). Road Extraction by Deep Residual U-Net. *IEEE Geoscience and Remote Sensing Letters*, **15**(5), 749-753.

[5] Yu, F., & Koltun, V. (2016). Multi-Scale Context Aggregation by Dilated Convolutions. In *International Conference on Learning Representations (ICLR).*

[6] Milletari, F., Navab, N., & Ahmadi, S. A. (2016). V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation. In *2016 Fourth International Conference on 3D Vision (3DV)* (pp. 565-571). IEEE.

[7] Ghorbanzadeh, O., Blaschke, T., Gholamnia, K., Meena, S. R., Tiede, D., & Aryal, J. (2019). Evaluation of Different Machine Learning Methods and Deep-Learning-Based Convolutional Neural Networks for Landslide Detection. *Remote Sensing*, **11**(2), 196.