Automatic Landslide Mapping with Interpretable Attention-based Convolutional Neural Networks Using Remote Sensing Data

Mulabbi, A.,[1, 4, 5] Danoedoro, P.[2 *] and Samodra, G.[3]
[1]Faculty of Geography, Universitas Gadjah Mada, Yogyakarta, Indonesia
[2]Remote Sensing Laboratory, Faculty of Geography, Universitas Gadjah Mada, Sekip Utara Jalan Kaliurang, Bulaksumur, Yogyakarta, 55281, Indonesia, E-mail: pdanoedoro@ugm.ac.id*
[3]Department of Environmental Geography, Faculty of Geography, Universitas Gadjah Mada, Sekip Utara Jalan Kaliurang, Bulaksumur, Yogyakarta, 55281, Indonesia, E-mail: guruh.samodra@ugm.ac.id
[4]Faculty of Education, Muni University, P.O. Box 725, Arua, Uganda, E-mail: a.mulabbi@muni.ac.ug
[5]School of Education, Uganda Christian University, Besania Hill, Mukono, Uganda
**Corresponding Author*

DOI: https://doi.org/10.52939/ijg.v21i7.4321

Abstract
Landslide mapping plays a vital role in disaster management by providing essential information that can help decision making on mitigation and early warning strategies. However, existing automated methods often lack interpretability and miss crucial details, which limit their practical utility. This study addresses these limitations by introducing a novel Spatial Attention U-Net that leverages human visual attention to improve landslide detection and interpretability. Our proposed method integrates spatial attention modules throughout the U-Net's encoder and decoder paths, enabling the model to focus on critical image features for landslide identification. The model is trained and evaluated using a combination of high-resolution Pleiades RGB imagery, Brightness Index, and slope data. The model's performance was evaluated using the F-1 score, precision, recall, and intersection over Union (IoU). The findings demonstrate that the Spatial Attention U-Net outperforms baseline models (U-Net, Squeeze-and-Excitation U-Net, and Channel-wise Attention), achieving F-1 scores of 73% and 79% on the testing and benchmark datasets, respectively. When applied to the inference/hold-out area, all the attention-based models outperformed the standard U-net, missing only three landslide events compared to five missed by the baseline model. Furthermore, the saliency maps reveal that the models focus on diverse regions of saliency, including edges, textures, tone, and brightness. The spatial attention U-net primarily highlights landslide edges (terrain discontinuities), while the baseline models use a mix of edges, texture, tone, and brightness. The results also indicate that dual-path attention does not lead to significant improvement in model accuracy. This approach offers a powerful tool for rapid and automated landslide mapping, indicating areas of saliency that can aid data annotation process by paying more attention to landslide object boundaries. The model interpretability further facilitates the creation of landslide inventories, especially in regions with limited ground truth data.

Keywords: Interpretable Model, Landslide Mapping, Landslide Detection, Saliency Maps, Spatial Attention

1. Introduction

Landslides pose a significant and growing threat to lives, livelihoods, and infrastructure worldwide. Triggered by a combination of natural hazards, including extreme weather, earthquakes, and tsunamis, as well as human activities, landslides disproportionately impact mountainous regions [1][2][3] and [4]. Southeast Asia, particularly vulnerable due to tectonic activity and monsoon climate, has suffered substantial losses. Since the 1990s, the region has accounted for over 60% of the estimated USD 4.5 billion in global economic damages from landslides [5]. The 2017 tropical cyclone that devastated Java Island, Indonesia, highlights the catastrophic consequences of these events [6]. Indonesia, consistently ranked among the top three countries for landslide fatalities [7][8] and [9], was severely affected in Pacitan regency, where hundreds of landslides and floods occurred [6][10] and [11].

To mitigate the devastating effects of landslides, comprehensive landslide inventories are essential. These data-driven resources inform disaster response planning, land-use management, risk assessment, and early warning systems [12].

Landslide inventories provide critical information on the location, timing, size, type, and volume of past events [13] and [14]. These datasets support disaster monitoring, susceptibility mapping, and emergency planning [15]. Given the rapid onset and severe consequences of landslides, timely information is crucial for effective rescue and recovery in affected areas. Remote sensing and advanced data analysis, including deep learning, offer powerful tools for rapidly assessing the landslide impacts. These technologies enable accurate landslide mapping, supporting rapid response and long-term recovery. Deep learning methods, such as semantic segmentation and attention-based models, effectively extract relevant information from large datasets and are also useful in data-limited areas.

Traditionally, landslide inventories have relied on ground surveys and manual image interpretation. While remote sensing offers large-scale mapping using satellite technology [16], challenges persist in complex terrains, particularly tropical regions with dense vegetation and frequent cloud cover [17]. Moreover, existing methods, including deep learning models, often lack transparency and face limitations due to data scarcity, computational constraints, and terrain variability [4]. To overcome these limitations, a novel approach is needed. While data fusion can improve results, it is computationally intensive. Therefore, human expertise and advanced machine-learning techniques are essential. Attention-based interpretable models can significantly enhance landslide mapping and risk assessment by detecting small landslides across vast areas with minimal data and computational requirements. This represents a significant step towards landslide management in highly susceptible regions.

Convolutional Neural Networks (CNNs) have demonstrated significant potential for landslide mapping through the analysis of remotely sensed data, often surpassing traditional methods [18][19] and [20]. CNNs, known for their powerful feature learning in image tasks, outperform manual feature design methods [21]. However, their use in remote sensing for landslide detection is still relatively new. While studies have explored various CNN architectures for both single-event and multi-temporal landslide detection, challenges persist [3][4] and [22], using the DNN backbone alone often fails to distinguish well between landslides from other changed objects. Study [22] proposed a model that uses the Efficient Hybrid Attentional Atrous Convolution to improve landslide detection by extracting high-resolution feature maps. In [4], a generalised CNN was used to map landslides based on pre- and post-event data, while [3] applied a deep transfer learning strategy to create a multitemporal landslide inventory. The neural networks used in these studies mainly consisted of convolutional, pooling, and fully connected layers to identify landslide attributes. However, most models apply attention only in the encoder, with limited focus on refined spatial decoding, deemed important to preserve spatial details. Also, many models are built based on complex architectures that lack transparency for practical application and interpretation.

A few studies have attempted to evaluate how spatial and dual-path attention can improve model performance, such as the attention-constrained neural network [23], and the dual-path attention [24]. Compared to the self-attention model, these approaches were found to improve accuracy by focusing only on the most important and relevant features. In [25], a comparison of two Siamese attention networks was done, while in [26] a simple multi-head attention module was used to improve detection. However, these models still suffer from parameter complexity, limited generalization, and a black-box nature. This limits their applicability in real-world disaster management applications, where model interpretability and approachability are crucial. Improving model interpretability and algorithm efficiency while maintaining precision is still a hurdle in landslide detection. Since landslide mapping requires analysis of both fine spatial details and overall context, the spatial attention U-net proposed in this study helps provide a more thorough understanding of surface conditions and preservation of spatial information, leading to better detection. Furthermore, its dual-path nature (encoder and decoder integration structure) effectively captures the complex shapes and textures of landslides, thus improving accuracy.

To tackle existing challenges in landslide mapping, we propose an interpretable spatial attention U-net that incorporates an attention module in both the encoder and decoder paths. This approach permits the model to focus on salient regions and avoid information loss in the deeper layers. It recalibrates feature maps in the decoder, which is useful for capturing small landslides, and increases the robustness of the model by suppressing false positives and increasing generalization across different terrains.

This research was guided by three objectives: (1) to create an interpretable dual path spatial attention (DPAS) U-Net with strong feature extraction, minimal parameters, and low complexity; (2) to assess the capability of the proposed model to map small and large landslides in complex tropical environments; and (3) to compare the spatial attention mechanism with other existing U-net architectures in terms of accuracy, efficiency, and interpretability. The rest of the article is organised as follows: Methods, Study Area and Materials (Sections 2), Results (Section 3), Discussion (Section 4), and Conclusion (Section 5).

2. Materials and Methods

2.1 Study Area

This study was based on a case study of Pacitan Regency. It is located in East Java Province, Indonesia, encompassing 1,390 square kilometres (km²) of land (Figure 1), and comprising 12 districts. The regency is nestled in Java's Southern Mountain Zone, with slopes ranging from 15 to 60 degrees and elevations varying from 50 to over 1,000 meters. Pacitan borders Ponorogo and Wonogiri regencies to the north, and the Indian Ocean to the south. The landscape is characterized by diverse relief and landforms, dominated by steep and hilly terrain. The Girindulu River serves as the area's primary drainage system, ultimately flowing into the Indian Ocean. Pacitan's diverse geology plays a significant role in its susceptibility to landslides. The region exhibits a variety of rock types, including volcanic breccia, lava tuff with sandstone and siltstone layers, basaltic pillow lava, and sedimentary rocks like sandstone, claystone, cherts, and limestone/karst [27]. These rocks constantly undergo intense weathering and erosion. Furthermore, Pacitan's location within an active tectonic zone leads to hydrothermal alteration, which weakens rocks and increases landslide risk [28]. The tropical monsoon climate brings high rainfall, with some months receiving up to 500 mm. The region is also prone to occasional cyclones, such as Cyclone Cempaka in 2017. The area experiences two distinct seasons: wet (October-April) and dry (May-September). Extreme climate events such as El Niño Southern Oscillation (ENSO) are a common feature and significantly influence monsoon patterns. Pacitan's population has grown substantially, leading to the expansion of settlements and farmlands. This land-use change disrupts natural landscapes and may increase landslide risk.

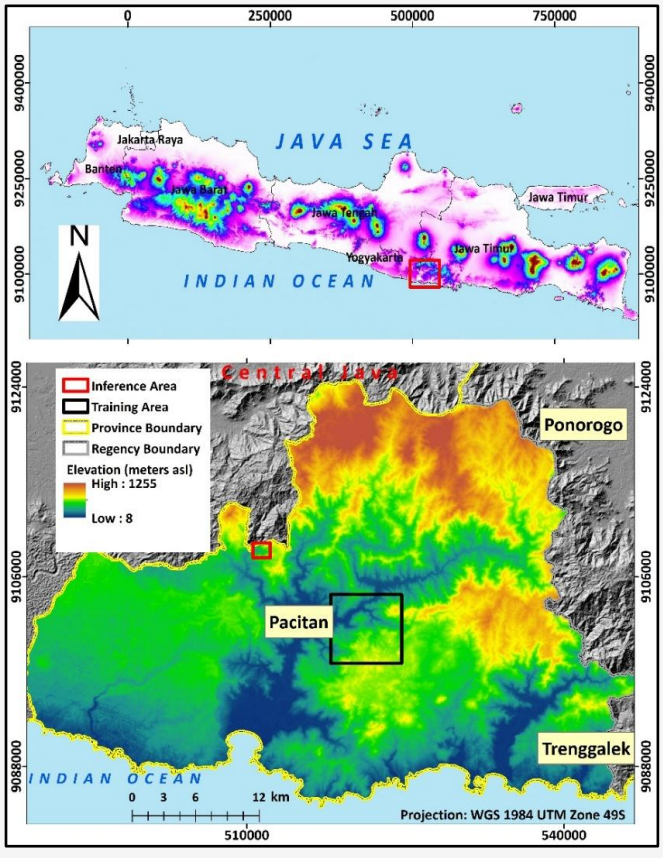

Figure 1: Pacitan regency, East Java: Training and inference sites

Given its unique geological makeup, climate, and population growth, further landslide research in Pacitan is essential. Such studies will help develop effective mitigation measures to safeguard lives and infrastructure. For training the model and testing its generalization ability, two sites were selected from the regency (Figure 1): a training site (for input data and predictions) and an inference site or image (a separate region to test generalization).

2.2 Data and Materials

This study used two datasets: the high-resolution (0.5 meters) Pleiades-1A satellite imagery from the Indonesian Space Agency (LAPAN) and the Bijie landslide dataset for validation**.** The image includes four bands: red/B2 (620 -700), green /B1 (510 – 590), blue /B0 (450 – 530), and near-infrared /B3 (775 – 915). The image was acquired on March 4, 2018, four months after the Tropical Cyclone Cempaka and the accompanying landslides. The image was pre-processed to ensure geometric and radiometric fidelity. This included georeferencing to establish real-world coordinates and image sharpening to enhance detailed features crucial for landslide detection. The processed image was clipped to match the designated study areas in Pacitan Regency. The Pleiades 1A image was preferred in this study for its very high spatial resolution, which enables accurate detection of even small landslides that dominate the study area. The Pleiades satellite also offers a frequent revisit cycle and multispectral capabilities for detecting soil exposure, vegetation loss, and debris flow paths. As such, this image is deemed suitable for emergency response planning and damage assessment. Other comparable datasets include Plane Scope, World View, SPOT, and other high-resolution images, though they typically require purchase. The Pleiades image used in this study was provided at no cost by LAPAN.

The Bijie landslide dataset was used for model validation as a standard benchmark to test model performance. This is an open-source dataset comprising remote sensing images from the city of Bijie, China. It was curated by researchers [29] from the University of Wuhan and made available online (http://gpcv.whu.edu.cn/data/Bijie_pages). It comprises 770 landslide image patches, 2,300 non-landslide patches, and corresponding DEM image tiles. This dataset was preferred for its high resolution, high quality, and completeness, including the Digital Elevation Model (DEM) layer, which is necessary for this study. The three image bands, the slope layer derived from the corresponding DEM, and the brightness index layers were all used to test the generalisation ability of the models.

Additionally, a DEM from the ALOS Palsar sensor was obtained from a public repository **(**https://search.asf.alaska.edu/**).** This DEM provided critical information on slope, which was used as a conditioning factor that helps the model distinguish terrain discontinuities between landslides and non-landslides. To identify landslide-prone areas, a brightness index band was calculated from the satellite image. The brightness index and slope raster were then combined with the original Pleiades 1A image, creating a new five-band image input dataset. Finally, all bands were rescaled to match the Pleiades image's high resolution, allowing for the detection of even small landslides prevalent in Pacitan Regency.

While satellite imagery is vital for landslide detection, supervised deep learning models also require ground truth data for training. A pre-existing point-based landslide inventory was available in this study. Although not suitable for directly training the U-Net model, the points were valuable in pinpointing landslide locations. These points were used to identify specific landslide instances within the satellite imagery. The identified landslide areas were then digitized into polygons, forming a crucial training dataset in the form of a shapefile.

2.3 Methods

This study evaluated the effectiveness of spatial attention and an interpretable U-Net architecture for automated landslide segmentation. Its accuracy and performance were compared with three baseline U-Net variants: the standard U-Net and two versions enhanced with distinct attention mechanisms (Channel-wise and SE). These attention mechanisms direct the model to focus on important visual features, which may help improve the model's accuracy. High-resolution satellite images, brightness index, slope data, and corresponding landslide masks were used to train and test the models. The analysis was performed on landslide objects, defined as groups of pixels. This analysis unit was chosen due to the high resolution of the image data used to extract landslide masks. To assess generalizability, two inference datasets were used: one from a separate area in Pacitan Regency and another from the Bijie landslide data. The best-trained models were used to predict landslides on the inference datasets. Model interpretation was done using saliency maps that visualize image regions influencing model predictions. This technique offers insights into the model's decision-making process, enabling domain experts to understand the underlying rationale behind landslide detection [30] and [31].

2.4 Preparation of Training Data

Supervised deep learning necessitates high-quality training data. In this study, satellite imagery and corresponding landslide labels (ground truth shapefile) were used as input data for the CNN models. To accommodate the CNN architecture, which processes data in smaller segments, both images and labels were divided into 128x128 pixel patches. This patch size was selected to optimize computational and memory efficiency. Each image patch included multiple spectral bands, including slope and brightness index, to enrich model's input. Landslide polygons, which include initiation, sliding, and deposit zones, were rasterized into binary masks aligned precisely with image patches, ensuring accurate training data. Data augmentation techniques were applied to the training dataset to improve model generalizability and mitigate data imbalance. Patches without landslide information (zero-valued) were excluded to improve data quality. Horizontal and vertical flipping, Gaussian blurring (σ = 0, 3), sharpening (α = 0, 1), shearing (−20, 20 degrees), and rotation (45 and 90 degrees) were applied to generate diverse training samples. These augmentations aimed to simulate realistic landslide patches with varying orientations and spectral characteristics. Different augmentation combinations were systematically conducted to optimize model performance. The augmented patches/dataset had the same specifications (resolution, number of bands, and patch size) as the original training data. The complete procedure from data preparation to model generalization testing is summarised in Figure 2.

2.5 Convolutional Neural Network Model Architecture and Training

The U-Net architecture was first proposed by [32] and is renowned for its effectiveness in segmenting images, especially medical images. This architecture was adopted as the backbone of this study. It comprises a contracting path for feature extraction and an expanding path for pixel-wise prediction. Three attention mechanisms (channel-wise, spatial, and SE) were added to the U-Net architecture specifically to address its limitations in handling imbalanced datasets and to improve segmentation accuracy [4]. These attention blocks improve the model's ability to focus on relevant image features [23]. The SE and Channel-wise attention blocks were integrated only in the decoder paths of the U-Net. Specifically, the SE and channel-wise attention blocks were integrated after the final convolution layer of the encoder path and at every skip connection point within the decoder path.

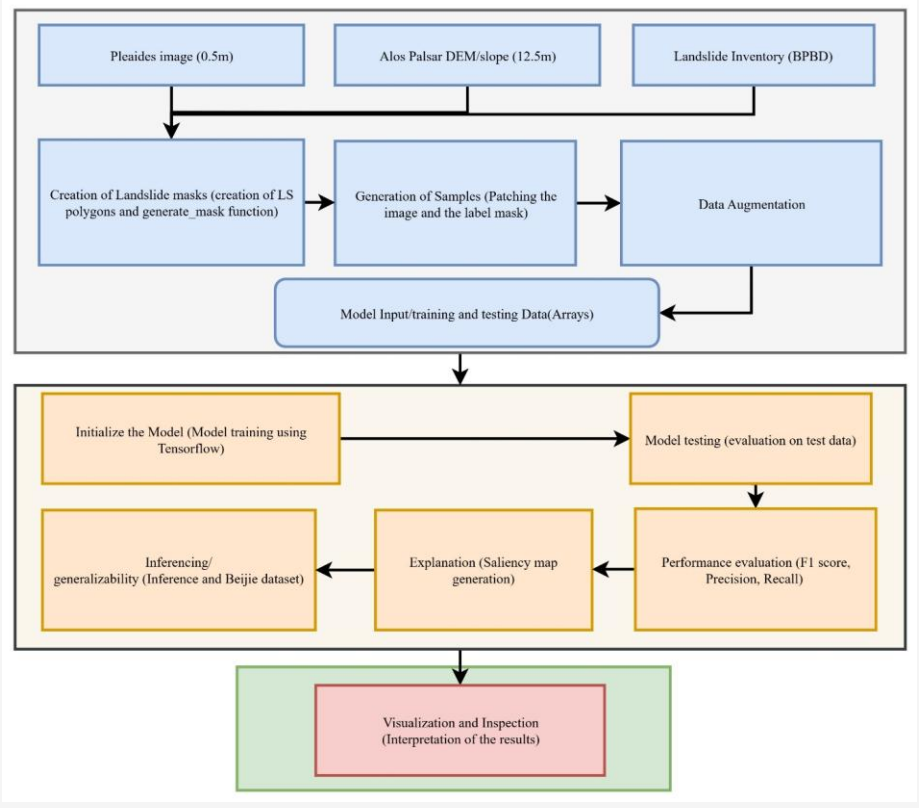

Figure 2: The conceptual flow of the methods adopted in this study (The procedure applies to all the models)

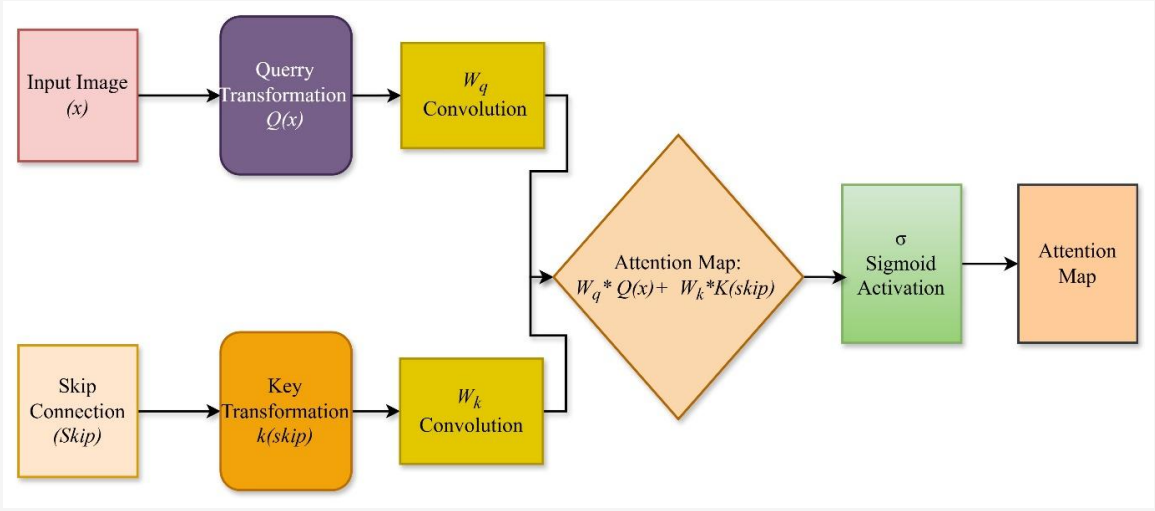

Figure 3: Schematic diagram of the proposed Spatial Attention mechanism

Spatial attention modules were added to both the encoder and decoder paths of the U-Net architecture. In the decoder section, attention blocks were applied at each stage to improve the skip connections from the encoder. This block uses self-attention to calculate scores through dot-product attention between the up-sampled decoder features (query) and the corresponding encoder features (key/value). The attention-weighted output was then applied to the skip connection and combined before the next convolutional layers. Spatial attention modules were placed after every max- pooling layer in the encoder and after every up-sampling operation in the decoder. This setup helps the model focus on important spatial areas during the merging of encoder and decoder features and also improves feature selection during reconstruction [24] and [33]. In the Dual-path attention U-Net model, the encoder captures context from the input image by down-sampling the image, while the decoder up-samples the encoder output to create a segmentation map.

The Spatial Attention Module (Figure 3) helps convolutional neural networks focus on specific parts of an image. It creates a map that highlights the most important areas based on how different parts of the image relate to each other. The spatial attention mechanism employed in this research is described by Equation 1 below, adopted from [34], which illustrates how spatial attention is calculated. The query and key features are transformed by their respective weight matrices and then summed. The sigmoid activation function constrains the output between 0 and 1, thereby representing the relative importance of each spatial location. This was applied to both the encoder and decoder paths of the model.

$$A(x, skip) = \sigma(W_q Q(x) + W_k K(skip))$$

Equation 1

Where:

$A(x, skip)$ is the Attention map: This represents the final attention map with values between 0 and 1.

Σ is the sigmoid activation function, ensuring the output is between 0 and 1.

W_q, W_k are the weight matrices for the query *(Q)* and key *(K)* convolutional layers, respectively.

$Q(x)$, $K(skip)$ are temporary maps created within the model. They're generated from the input image *(x)* and the skip connection (additional information from earlier processing) using small filter layers.

A ReLU activation function is applied to add non-linearity and improve the model's ability to learn complex patterns.

By focusing on specific image regions based on their spatial coordinates, this module improves the model's ability to highlight critical components. This is achieved by transforming the input and skip connection features with weight matrices, followed by summation and sigmoid activation to generate attention weights (Figure 3). These weights, ranging from 0 to 1, quantify the importance of each spatial location. Given its emphasis on spatial context, this mechanism is particularly effective for tasks requiring precise localization, such as image segmentation, object detection, and classification, where identifying and prioritizing crucial image elements is essential. During model training, Adam optimizer was used to optimize training and the ReLU activation function was used to help account for non-linearity in the input data.

Finally, a 1x1 convolutional layer with sigmoid activation was used to generate landslide probability maps.

A threshold of 0.5 was then applied to produce binary segmentation output. The 0.5 threshold was preferred because in binary classification, it represents the midpoint between classes 0 and 1, especially when the classes are balanced [35]. To verify the suitability of the 0.5 threshold, a histogram of the prediction probabilities was created following the method in [36] to observe the value distribution. The histogram indicated a clear separation between classes 0 and 1, making the 0.5 threshold an optimal choice.

To achieve optimal performance, this study employed a trial-and-error approach for hyperparameter tuning, aiming for a combination that minimizes loss. The chosen hyperparameters included 32 filters, a batch size of 16, a learning rate of 0.00005, and 150 epochs. To monitor the training and prevent overfitting, a set of call-backs was implemented, including early stopping, learning rate scheduling, and model checkpointing. Additionally, a validation split of 30% was applied to the training data to create a dedicated validation set for model evaluation. This validation step is crucial for building a robust U-Net model, as it offers insights into performance, helps prevent overfitting, and helps the final model generalize to unseen data. This research utilized Python within a Google Colab Pro Jupyter Notebook environment for model development, training, testing, and validation. GDAL and ArcMap were employed for GIS processing, while Keras and TensorFlow powered deep learning tasks. To address class imbalance and focus on challenging samples, a combined Tversky index (*TVC*) and Focal Tversky loss function (*TVF*) was implemented. This loss function combines the strengths of both metrics, making it suitable for segmentation tasks. The Tversky index (*TVI*), Tversky loss *(TVL),* Focal Tversky loss function (*TVF*), and combined Tversky index (*TVC*) are determined from Equations 2 to 5 [37]:

$$TVI = \frac{TP}{TP + \alpha \cdot FN + \beta \cdot FP}$$

Equation 2

Where:

TP is True Positives that represent the sum of the correctly predicted positive values.

FN is False Negatives that represent the sum of the actual positives that were predicted as negative.

FP is False Positives that represent the sum of the negatives that were predicted as positives.

α is the weights that control the penalties for false negatives.

β is the weights that control the penalties for false positives, respectively.

$$TVL = 1 - TVI$$

Equation 3

$$TVF = (1 - TVI)^{\gamma}$$

Equation 4

Where:

γ is a focusing parameter that adjusts the rate at which easy examples are down-weighted.

$$TVC = TVL + TVF$$

Equation 5

The training was terminated if validation loss did not decrease for 30 consecutive epochs. The model with the lowest validation loss was subsequently selected for evaluation using both test and inference data.

2.7 Model Explanation and Validation Metrics

The trained U-Net models were evaluated on the test dataset comprising satellite image patches and corresponding landslide masks. *F1-score*, *precision*, and *recall* metrics were computed to assess model generalization. These metrics provide a more comprehensive performance evaluation than accuracy alone [38]. *Precision* is defined as the proportion of correctly identified landslides among all predicted landslides, while recall is the proportion of actual landslides correctly identified. The *F-1 score* balances *precision* and *recall,* offering a robust indicator of overall model performance. These metrics are preferred in this study because the landslide detection datasets are imbalanced, with non-landslide pixels far outnumbering landslide pixels. Therefore, *the F-1 score*, as the harmonic mean of *recall* and *precision*, helps balance detection accuracy between landslide and non-landslide classes [39]. Equations 6 to 10 are used to calculate the model performance metrics.

$$Precision = \frac{TP}{TP+FP}$$

Equation 6

$$Recall = \frac{TP}{TP+FN}$$

Equation 7

$$F\text{-}1\ Score = \frac{2Precision \times Recall}{Precision + Recall}$$

Equation 8

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Equation 9

$$IoU = \frac{Area\ of\ union}{Area\ of\ overlap}$$

Equation 10

Where:

IoU = Intersection over Union

Model generalization was evaluated by applying the trained model to make predictions on the inference datasets. First, the Bijie dataset and the F-1 score were used to assess performance. Second, the model was applied to the inference region within Pacitan regency, where predictions were made and visualized using Contextily and compared to a high-resolution, point-based landslide inventory.

2.8 Model Explanation Using Saliency Maps

To understand the model's decision-making process, saliency maps were generated as part of the model output using the Vanilla gradient method proposed by [40]. This method computes the gradient of the model's average prediction score and considers the absolute mean across channels to highlight the most influential pixels. Gradient defines the sensitivity of the model's prediction to each input data pixel, implying that a higher gradient pixel means greater influence on the prediction. Saliency maps are generated by computing the gradient of the model's output for the input images (test data). The gradient thus indicates the extent to which a minor alteration in the value of each input image pixel would influence the model's output. These maps highlight the image regions most influential to the model's predictions. Brighter areas indicate greater importance. Saliency maps differ from ordinary maps, such as the Landslide susceptibility maps, due to their ability to represent saliency intensity levels of the most influential pixels in the input dataset. This offers additional information revealing the internal workings of the Deep learning model and can be used to assess the model's behaviour and even its geomorphic plausibility. Saliency maps improve landslide detection accuracy by aiding model refinement, especially in reducing bias, thereby improving performance and trust in the model's outputs [41]. Mathematically, according to [40] the saliency at pixel *i, j* is expressed in Equation 11:

$$Saliency(i, j) = \frac{1}{C} \sum_{c=1}^{C} \left| \frac{\partial L}{\partial I(i, j, c)} \right|$$

Equation 11

Where:

L is the loss function of the model.

I(i, j, c) is the pixel value at a specific location in the input image.

C is the number of input image channels.

∂L / ∂I (i, j, c) is the loss gradient at a specific pixel.

3. Results

3.1 Results of the Comparative Experiments

The proposed spatial attention U-Net was compared to the standard U-Net, SE block, and channel-wise attention using identical hyperparameters. Figure 4 shows the loss curves, indicating that the spatial attention U-net exhibited a more stable training pattern than the other models. Noticeable spikes in loss are evident in Figure 4 (b), (c), and (d), where the loss curves exhibit instability and fail to smooth out, unlike the spatial attention U-Net (Figure 4(a)). The SE U-Net achieved the lowest training loss, followed by standard U-Net, channel attention, and spatial attention variants. Despite this, the attention-based model demonstrated better overall landslide detection performance, particularly in capturing landslide edges and shapes. This suggests that attention mechanisms, while potentially influencing training dynamics, significantly improved the model's ability to identify landslide features.

Additionally, three metrics were used to further evaluate the models. The training and testing performance metrics are presented in Table 1. Based on the F1-score, the standard U-Net and the channel-wise attention models produced the highest training scores of 0.98, followed by the SE attention model at 0.97, and the spatial attention at 0.74. The models showed similar performance on the test dataset. The spatial attention U-Net performed slightly better than the other three models, achieving an F-1 score of 0.73, followed by the standard U-Net at 0.71, and both the channel-wise and SE attention models at 0.70 (Table 1). The spatial attention U-Net exhibited less variation between training and testing metrics than the other models. This suggests that the spatial attention U-Net is more robust, as the other models demonstrated a substantial discrepancy, indicative of overfitting.

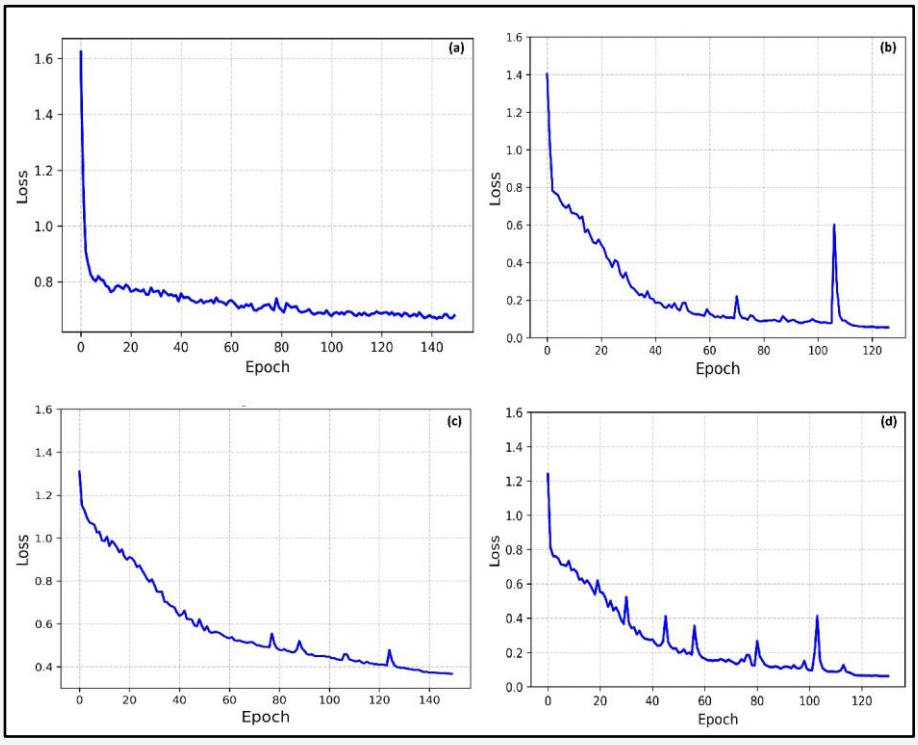

Figure 4: Training Loss for the four models:
(a) Spatial attention, (b) ordinary U-net, (c)channel-wise, (d) SE attention

Table 1: Training and testing performance metrics for the four model architectures

Model	Training			Testing accuracy				IoU
	Precision	Recall	F1-score	Precision	Recall	F1-score	Accuracy	
U-Net	0.99	0.96	0.98	0.78	0.66	0.71	0.89	0.53
U-Net with Channel att.	0.98	0.97	0.98	0.80	0.61	0.70	0.89	0.55
U-Net with spatial att.	0.77	0.71	0.74	0.72	0.74	0.73	0.90	0.58
U-Net with SE att.	0.96	0.98	0.97	0.80	0.64	0.70	0.89	0.54

**Remark: att. is attention*

Based on the precision score, the channel-wise and SE attention models outperformed the standard U-net and the spatial attention. In terms of IoU, which measures the overlap between predicted and actual or ground truth masks, the spatial attention U-net outscored the other U-net models by over 3 percentage points. The SE and channel-wise attention models demonstrated marginally higher recall than the standard U-Net, while spatial attention exhibited the lowest. Recall measures the model's ability to detect all actual landslides. All models exhibited false alarms, particularly for small, shallow landslides lacking distinct terrain and visual features. While the SE and spatial attention models struggled with shadowed landslides, the channel attention model generally outperformed the other models. The Spatial attention model excelled in capturing landslide morphologies but struggled with poorly defined boundaries (often due to inaccurate mask labelling), leading to lower recall and precision due to false negatives. Conversely, the channel attention model's high recall was offset by a high false positive rate, frequently misidentifying structures such as house roofs as landslides. The spatial attention U-Net was also able to discriminate between landslides and other objects, such as house roofs, which exhibited a similar brightness to landslide objects, as seen in Figure 5(b). These findings emphasize the need for a comprehensive evaluation using multiple metrics beyond recall.

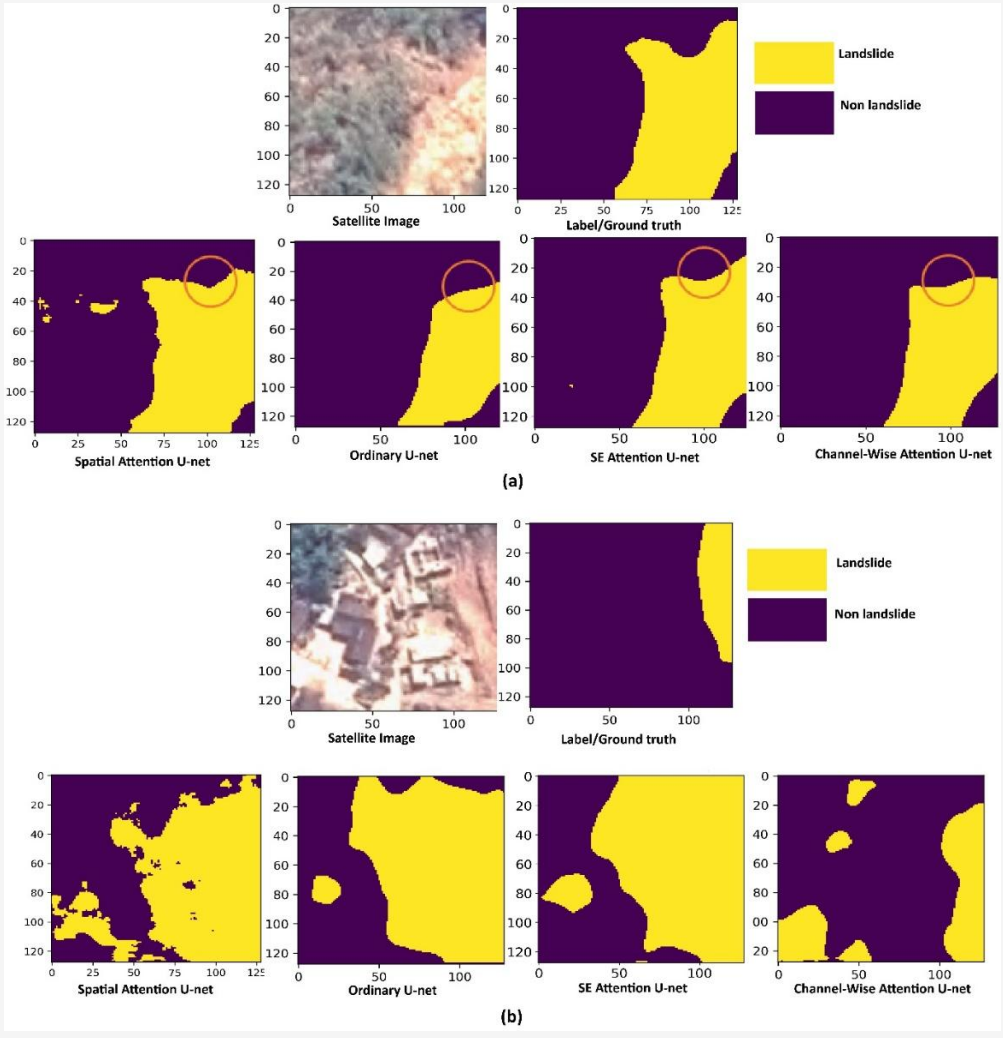

Figure 5: Model predictions on test data (a) without houses, (b) with houses

3.2 Predictions by the Models

After training, the model with the optimal weights and the lowest validation loss was employed to make predictions on the test dataset. Sample outputs are presented in Figure 5. All three models performed well on the sample image, clearly distinguishing the landslide objects from the non-landslide pixels. The spatial and SE attention models produced results closest to the ground-truth image, especially in preserving the landslide shape (Figure 5(a), where the red circles highlight the differences in precision). The red circles also indicate that the SE, channel-wise attention, and standard U-net models provide less detail and precision in capturing landslide shapes. Fine terrain discontinuities captured by the spatial attention U-net outside of the main landslide mass were missed by the other three models, indicating differences in precision. The spatial attention model also performed well in differentiating between house roofs and the mudflows/soil masses within the compound, even though the channel-wise attention model produced results closer to the label image/mask (Figure 5(b)). The models also performed at varying levels of precision in differentiating soil masses from house roofs (Figure 5(b)). Even when the ground truth label omitted soil masses within the housing complex, the spatial and channel-wise attention mechanisms refined the landslide boundaries. However, the varying roof materials in the sample image influenced the model's ability to differentiate between landslide and non-landslide objects/pixels. Notably, the spatial attention model provided better differentiation between dark-coloured house roofs (likely ceramic tiles) and bright coloured/shining house roofs (likely aluminium sheets). The SE and standard U-net models produced nearly identical results, failing to clearly distinguish landslides from house roofs in the predictions (Figure 5(b)).

3.3 Model Performance on Inference Dataset

To assess model generalization, all four U-Net models were tested on the Bijie landslide dataset and an inference area in Punungu Subdistrict, Pacitan Regency. Predictions were made using the models with the lowest validation loss, and the results were then compared to the ground truth/mask data. An F-1 score of 79% and precision score of 83% (Figure 6) revealed that the spatial attention U-net performed slightly better than the other models on the Bijie dataset. On the inference image from Pacitan Regency, all models performed well. However, visual analysis of the results indicated that all attention-based U-net models outperformed the standard U-net model, missing out only three landslide points based on the ground truth data, compared to five missed by the standard model (Figure 7).

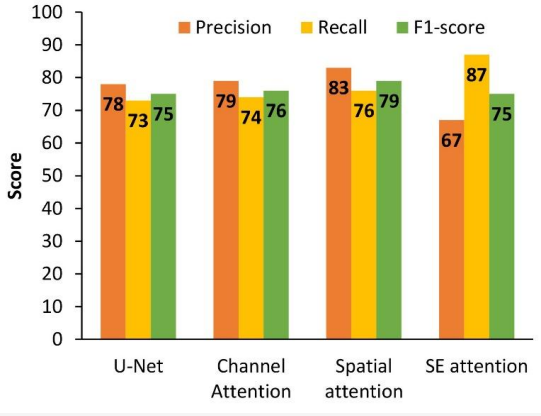

Figure 6: Model validation scores on the open landslide dataset

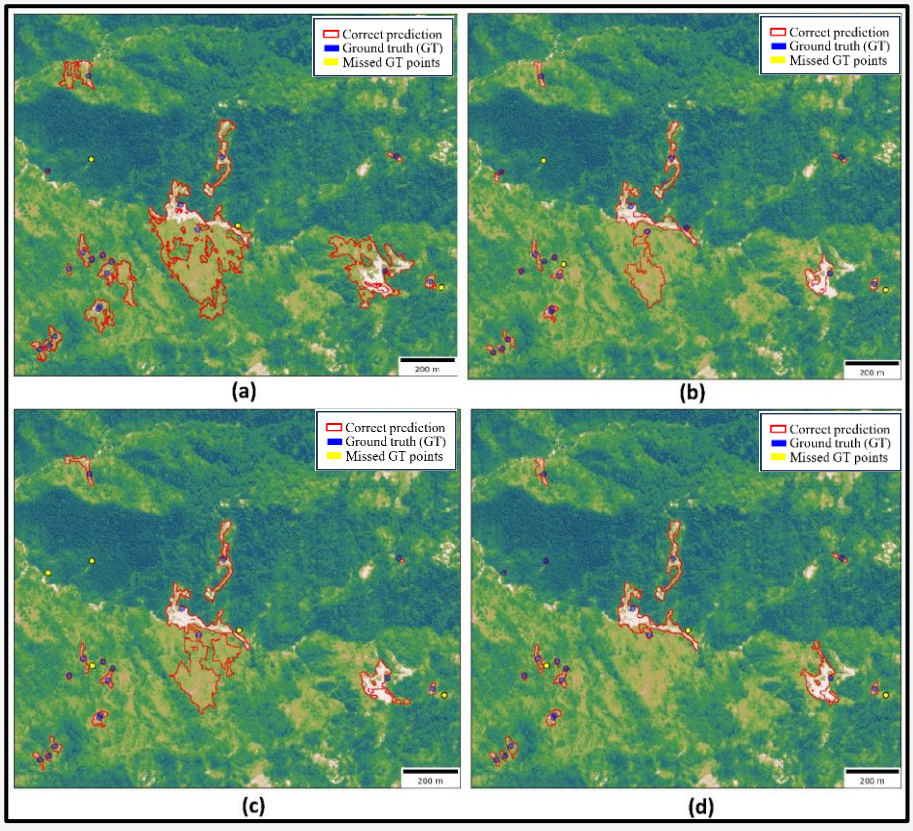

Figure 7: Ground truth and landslide segmentation using different models (a) spatial attention U-net, (b) SE attention U-net, (c) Standard U-net, (d) Channel-wise attention U-net

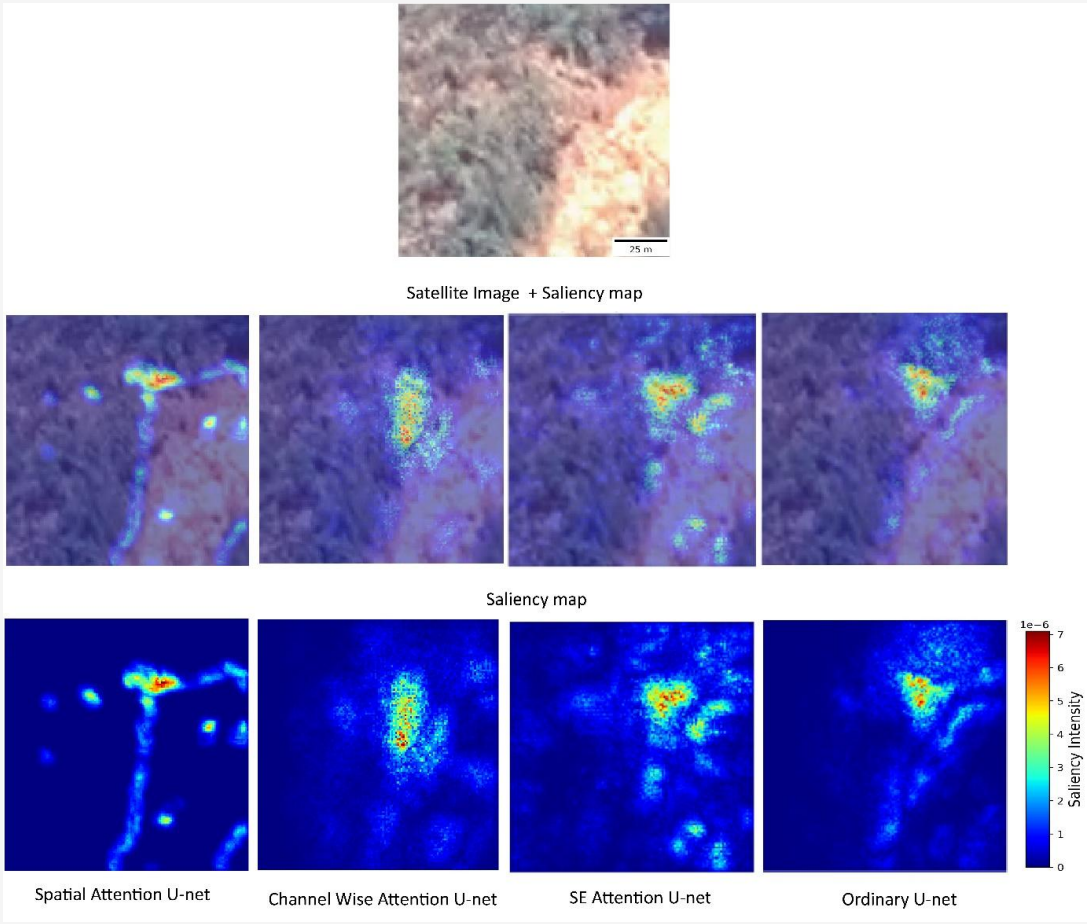

Figure 8: Saliency Maps of the four CNN models

3.4 Model Interpretation

The saliency maps in Figure 8 highlight image regions most influential to the model's predictions, with brighter areas indicating greater importance. Figure 8 shows that the spatial attention model primarily focuses on edges, while the other models also consider brightness and texture. All models highlighted the same image section with the highest saliency intensity, indicating some level of similarity in their attention mechanism. Closer inspection reveals that the high saliency intensity region corresponds to a sudden slope break or terrain discontinuity between landslide and non-landslide areas. This implies that terrain discontinuities serve as salient features influencing the model's decision-making. In addition to terrain discontinuities, the channel-wise, SE, and standard U-net models also exhibit other saliency features, as seen from the uneven distribution of saliency intensity across the entire landslide object. Saliency maps offer valuable insights into model behaviour by identifying which image features most influence the decision mechanism of the models. Other than object edges, the models also rely on texture, brightness, colour, and tone when making predictions.

4. Discussion

4.1 Comparison of the Model's Training and Testing Performance

Accurate and timely landslide mapping is crucial for effective disaster management. Deep learning models offer potential for automation and efficiency. However, introducing attention mechanisms increases model complexity, as measured by the number of parameters. While attention-based models exhibited more stable training (Figure 4), they also tended to overfit (except for the spatial attention U-Net) as indicated by higher loss values. This overfitting arises from the models focusing excessively on specific training data regions. [42] and [43]. Consequently, attention-based models should be applied with caution, and their results should be interpreted critically.

The loss curves indicate unstable training for the SE, channel-wise, and standard U-Net models, with sharp spikes after epoch 100. This behaviour may result from the use of channel attention in both channel-wise and SE models, that has been found to experience training instability caused by abrupt shifts in channel importance, especially without proper regulation [44]. The instability in the standard U-Net training loss could be attributed to its susceptibility to overfitting and local minima [45]. Within contrast, the spatial attention U-Net model exhibits a smoother training loss curve, likely due to its focus on informative pixels without necessarily suppressing other input image channels [43].

The spatial attention U-Net demonstrated marginally better performance, with an F-1 score of 0.73 and an IoU score of 58%, compared to the standard, SE, and channel-wise U-Net models. This improvement is attributed to the attention mechanism, which enables the model to focus on crucial features. Although the improvement is not significant enough to qualify the spatial attention model as the most effective for landslide delineation from satellite images, its superior F1, IoU, and precision scores on the Bijie dataset indicate its effectiveness for landslide mapping. The performance gains in all attention-based models, especially on the inference image, aligns with previous research highlighting the benefits of attention mechanisms in U-Net architectures [34][46] and [47].

As observed from the results, all models produced both false positives and false negatives. False positives in landslide delineation using Deep learning models have been linked to dataset uncertainty [25], annotation errors (missed details), shape and texture misinterpretation, and other environmental and model-related reasons. For example, the spatial attention U-net produced false positives, possibly due to over-reliance on terrain discontinuities to differentiate between landslide and non-landslide objects where similar saliency is detected in the image (Figure 5). As noted by [48] and confirmed in this study, annotation might introduce uncertainty in the data, which eventually reduces detection accuracy. Figure 5 corroborates these findings, showing the spatial attention U-net refining the landslide boundaries based on salient features. The robust performance score of the standard U-net on the training and the Bijie datasets indicates that greater model complexity does not always lead to improved detection accuracy. Accordingly, the models need to be designed with appropriate complexity and trained and evaluated using suitable loss functions and performance metrics to enable rapid and accurate landslide detection, crucial for emergency response and disaster monitoring.

Although the input dataset included five bands, the results may not reflect the importance of slope and Brightness index in the models. This is because the model setup does not account for the saliency contribution of individual bands. Nonetheless, occlusion intensity analysis showed that all the bands contributed to the decision process, albeit at varying levels. However, because the results are not significantly better than those from other studies using only the usual RGB bands, additional indices like slope and brightness may be less impactful. This finding contrasts with [49], who noted that brightness and differential brightness indices improved landslide detection when using a combination of optical and SAR images. Therefore, the number of input variables used may not be the only factor influencing model performance. As such, model accuracy also depends on the data context and characteristics, the nature of the landslide objects, and the model architecture [20].

4.2 Model Generalization Ability on the Inference and Bijie Datasets

This study evaluated model generalization using test data, inference images, and the Bijie data predictions [50]. The spatial attention U-Net outperformed the other models on the Bijie dataset, producing more consistent results in terms of precision and F-1 scores. The SE attention model showed better precision in the inference image, and along with the channel and spatial attention models, missed only three landslide objects when compared with the ground truth points. The improvement by the spatial attention model on the Bijie dataset may be attributed to the model's ability to capture spatial relationships and dependencies within images [51], enhance feature representation [43] and improve localization [45]. Unlike channel attention, spatial attention focuses on the spatial configuration of features [52].

Although all models demonstrated reasonable performance in identifying landslide instances, false positives were prevalent particularly in open areas, with some models exaggerating the landslide polygons. This is likely attributable to the 128 x 128 patch size [53], which may be insufficient to highlight contextual information during prediction [25]. Additionally, the low landslide density in the area, where most image patches contain only one or even half landslide masks, may hinder model learning, leading to diminished performance on the inference image. The spatial attention U-Net performed well on the Bijie dataset but did not generalize as effectively to the inference image.

This variation in performance may be attributable to the nature of the data; in the Bijie dataset, each image patch covers the entire extent of the landslide, offering more contextual information for more accurate prediction. On the other hand, the inference area has a mix of both large and small slides. The use of 128 x 128 patch size may distort contextual information, which is valuable for Deep Learning models to make accurate predictions [54]. Additionally, complex terrain features resembling landslides, such as rocky areas, dry riverbeds, or cleared land/rice fields, can also lead to these errors [55]. The quality of data annotation could also be a reason for this performance discrepancy, as previously noted by [48]. Nevertheless, the spatial attention U-net model demonstrates greater robustness and generalizability for landslide extraction, with a higher true identification rate (ground truth landslides vs. predicted) compared to the standard U-net model. This supports the value of incorporating attention mechanisms in U-net models, as previously noted by [29]. Accordingly, the patch size should be determined based on the context of the study area, the size of the landslides, and computational resources.

4.3 Model Interpretation

Deep learning models, similar to humans, selectively focus on image features during decision-making [40] and [41]. Understanding these focal points is crucial. Saliency maps visualize model attention, and Figure 8 reveals that the spatial attention U-Net model effectively prioritizes edges, which represent terrain discontinuities leading to more refined landslide shape and boundary delineation. All models use terrain discontinuities such as scarps, indicative of landslide initiation zones, as salient regions. However, their focus varies: channel attention focusing on edges and textures, standard U-Net on texture and brightness, and SE block on textures. Geomorphologically, these salient features (edges, texture, brightness) represent terrain discontinuities like scarps, especially along sharp slope transitions. These geomorphological features often appear as edges in DEMs and satellite images, as indicated by the saliency maps in this study, and always point to areas of head scarps or unstable terrain slopes. This is supported by [12], who noted that sudden slope changes are reliable signs for the detection of both active and incipient landslides. Also, [56] in their study on landslide susceptibility reported that Knick lines and slope transitions, which are classic geomorphological edge indicators associated with landslide initiation areas, helped improve the CNN model performance.

The diverse saliency attention patterns influence model performance, as seen in the varying landslide object representations. This variation in saliency has also been linked to false positive predictions, especially for the spatial attention U-net model. The standard U-net, the channel attention, and SE attention produced more precise results in both training and testing (Table 1), possibly due to their broader focus on multiple salient areas. Despite their sensitivity to terrain discontinuities, these models did not erroneously detect false positives similar to the spatial attention model, likely due to the existence of terrain discontinuities outside the main landslide area (Figure 5). Therefore, Saliency maps offer valuable insights into model behaviour, a previously overlooked aspect in landslide mapping, and can help refine the model design architecture to improve decision accuracy [41]. It is also evident that while saliency maps effectively highlight feature importance within training data, their predictive power on unseen images is limited. Despite consistent performance on inference, low accuracy indicates a mismatch between salient features and actual landslide regions. Brightness-based false positives, such as confusing bare land or riverbeds with landslides, further underscore this discrepancy. This could be because saliency attention was used only for interpretation, not during model training. Alternatively, this may result from attention drift across training, testing, and inference predictions. As noted by [57], attention mechanisms may focus on spurious features correlated with landslide labels and fail to learn causally relevant features. Integrating saliency maps into the model architecture could improve the model's landslide detection accuracy [58], ultimately transforming the saliency maps from interpretive tools into guidance or supervision mechanisms.

The results of this study demonstrate the significant potential of the spatial attention mechanism in U-Net models for various practical applications. The model's accuracy and focus on specific image attributes [47] make it a valuable tool for disaster response planning and management. Its adaptability across diverse terrains supports its use for global disaster monitoring efforts. Beyond landslide detection, the model can support analysis of other remote sensing data and contribute to early warning system design. As Figure 9 illustrates, rapid vegetation growth can obscure landslide scars, emphasizing the need for swift mapping. This aligns with previous research highlighting the urgency of mapping event-based landslides, especially in tropical regions [3] and [4].

Figure 9: Field photos and Google Earth images show how the vegetation has quickly regrown after landslide occurrence (a) 2018, (b) 2023, (c) 2024

5. Conclusions

This study demonstrates the effectiveness of the spatial attention U-Net in improving landslide detection accuracy compared to traditional U-Net architectures. The model's ability to focus on critical image features led to an improved F-1 score of 73%, surpassing the performance of other models by approximately 2 percentage points. The results also show that, though associated with false positives, the model generalizes well across domains, performing consistently on both the Bijie dataset and the inference image, suggesting its applicability in diverse geographic regions in the world. The results also confirm that attention mechanisms improve detection accuracy of the models, with all the attention-based models performing well on training, testing, inference, and the Bijie datasets. Additionally, the proposed model emphasizes salient features such as edges and boundaries, which correspond to terrain discontinuities like scarps.

Overall, all the models used in this study could be useful for landslide detection from satellite imagery, given that they all performed well across training, testing, and inference data. The spatial attention U-net model did not significantly outperform the other models, but showed marginal improvement in the evaluation metric scores. Therefore, the dual-path encoding approach may not offer a substantial advantage in landslide detection accuracy, as performance differences between the dual-path and single-path attention models were minimal. The robust performance of the standard U-Net model suggests that increased model complexity does not necessarily lead to improved accuracy. Nonetheless, the proposed model shows potential for practical application in disaster management. Additionally, saliency maps help interpret model decisions and enhance usability; they also have potential as supervisory tools when integrated into model training architectures. We believe that this approach can be used to study landslide areas with conditions similar to those used in this study, both locally and globally.

6. Suggestions and Limitations

This study underscores the importance of high-quality training data in landslide detection using deep learning models. Although a substantial, multi-variable dataset was used, our models encountered limitations during upscaling. Inaccurate sampling and imbalanced data distribution reduced the model's predictive accuracy, especially for underrepresented landslide types. Moreover, memory constraints and the use of relatively older Digital Elevation Model (DEM) data, predating Cyclone Cempaka, contributed to a high false positive rate. The DEM resolution also determines whether slope layers influence the model's predictions. As noted by [59], slope has a greater impact on model performance when low-resolution DEMs are used. This may explain the minimal influence of slope layers in this study. Other challenges include the varying size of landslides, particularly the limited pixel representation of small ones, which hindered accurate detection, consistent with findings from previous studies. The high-resolution Pleiades image is more effective in mapping landslides in areas where many landslides exist, so that a clear distinction can easily be made between landslide objects and other land cover. This may contribute to the high number of false positives, with the model struggling to distinguish certain land cover types from landslide objects. Although the model showed promise in detecting smaller landslides, further

improvements are required, potentially through techniques like pyramid pooling. To improve landslide inventory generation, accurate ground truth labelling and rigorous field verification are essential to complement automated procedures. Addressing these limitations will help future studies improve landslide detection models and provide more reliable results.

7. Future Work Recommendations

Future research will need to incorporate additional variables, including image GLCM, spectral indices (NDVI, NDWI), and refined topographic features (hill shade, elevation, TRI, updated DEM) to improve model performance. Data fusion with other sensor types and post-processing procedures will also be explored to reduce false positives and improve overall performance metrics. Utilizing saliency maps not only as interpretation tools but also as supervision tools during model training could help reduce false positives and improve model accuracy. To deepen our understanding of model behaviour, we will experiment with more sophisticated attention mechanisms (multi-head, Convolutional Block Attention Module) and deeper network architectures. Beyond saliency maps, partial dependency plots will be utilized to analyse feature-prediction relationships, providing insights into the model's decision-making process to support model refinement.

Acknowledgment

The authors would like to thank the KNB (*Kemitraan Negara Berkembang*) scholarship for the opportunity and funding it provided to the first author in the form of tuition, books, and research fees. Special thanks also go to the Indonesian National Institute of Aeronautics and Space (LAPAN) for providing the satellite imagery used in this study and the Disaster Management Agency of Pacitan Regency (BPBD) for the landslide inventory database.

References

[1] Akosah, S., Gratchev, I., Kim, D. H. and Ohn, S. Y., (2024). Application of Artificial Intelligence and Remote Sensing for Landslide Detection and Prediction: Systematic Review. *Remote Sensing*, Vol. 16(16). https://doi.org/10.3390/rs16162947.

[2] Piralilou, S. T., Shahabi, H., Jarihani, B. and Ghorbanzadeh, O., (2019). Landslide Detection Using Multi-Scale Image Segmentation and Different Machine Learning Models in the Higher Himalayas. *Remote Sensing*, Vol. 11(21). https://doi.org/https://doi.org/10.3390/rs11212575.

[3] Bhuyan, K., Tanyaş, H., Nava, L., Puliero, S., Meena, S. R., Floris, M., van Westen, C. and Catani, F., (2023). Generating Multi-temporal Landslide Inventories through a General Deep Transfer Learning Strategy using HR EO Data. *Scientific Reports*, Vol. 13(1), 1–26. https://doi.org/10.1038/s41598-022-27352-y.

[4] Prakash, N., Manconi, A. and Loew, S., (2021). A New Strategy to Map Landslides with a Generalized Convolutional Neural Network. *Scientific Reports*, Vol. 11(1), 1–15. https://doi.org/10.1038/s41598-021-89015-8.

[5] CRED. (2021). Disaster Year in Review 2020 Global Trends and Perspectives. *Cred*, *May* (62), 2020–2021. [Online]. Available: https://www.preventionweb.net/publication/cred-crunch-issue-no-62-may-2021-year-review-2020-global-trends-and-perspectives. [Accessed: Mar 4, 2024].

[6] Samodra, G., Ngadisih, N., Malawani, M. N., Mardiatno, D., Cahyadi, A. and Nugroho, F. S., 2020(a). Frequency–magnitude of Landslides Affected by the 27–29 November 2017 Tropical Cyclone Cempaka in Pacitan, East Java. *Journal of Mountain Science*, Vol. 17(4), 773–786. https://doi.org/10.1007/s11629-019-5734-y.

[7] Chian, S. C. and Wilkinson, S. M., (2015). Feasibility of Remote Sensing for Multihazard Analysis of Landslides in Padang Pariaman during the 2009 Padang Earthquake. *Natural Hazards Review*, Vol. 16(1). https://doi.org/10.1061/(ASCE)NH.1527-6996.0000143.

[8] Putra, A. N., Nita, I., Jauhary, M. R. Al, Nurhutami, S. R. and Ismail, M. H., (2021). Landslide Risk Analysis on Agriculture area in Pacitan Regency in East Java Indonesia using Geospatial Techniques. *Environment and Natural Resources Journal*, Vol. 19(2), 141–152. https://doi.org/10.32526/ennrj/19/20201 67.

[9] Samodra, G., Chen, G., Sartohadi, J. and Kasama, K., (2018). Generating Landslide Inventory by Participatory Mapping: An example in Purwosari Area, Yogyakarta, Java. *Geomorphology*, Vol. 306, 306–313. https://doi.org/10.1016/j.geomorph.2015.07.035.

[10] Darminto, M. R., Widodo, A., Alfatinah, A. and Chu, H. J., (2021). High-Resolution Landslide Susceptibility Map Generation using Machine Learning (Case Study in Pacitan, Indonesia). *International Journal on Advanced Science, Engineering, and Information Technology*, Vol.

11(1), 369–379. https://doi.org/10.18517/ijaseit.11.1.11679.

[11] Khuc, T., Truong, X., Tran, V., Bui, D., Bui, D., Ha, H., Tran, T., Pham, T., and Yordanov, V. (2023). Comparison of Multi-Criteria Decision Making, Statistics, and Machine Learning Models for Landslide Susceptibility Mapping in Van Yen District, Yen Bai Province, Vietnam. *International Journal of Geoinformatics*, Vol. 19(7), 33–45. https://doi.org/10.52939/ijg.v19i7.2743.

[12] Guzzetti, F., Mondini, A. C., Cardinali, M., Fiorucci, F., Santangelo, M. and Chang, K. T., (2012). Landslide Inventory Maps: New Tools for an Old Problem. *Earth-Science Reviews*, Vol. 112(1–2), 42–66. https://doi.org/10.1016/j.earscirev.2012.02.001.

[13] Jacobs, L., Dewitte, O., Poesen, J., Delvaux, D., Thiery, W. and Kervyn, M., (2016). The Rwenzori Mountains, a Landslide-prone Region? *Landslides*, Vol. 13(3), 519–536. https://doi.org/10.1007/s10346-015-0582-5.

[14] Torgoev, I., Akylbek, U. B. and Chymyrov, A., (2022). UAV Survey for Landslide Hazard Assessment in the Former Min-Kush Uranium Processing Site. *International Journal of Geoinformatics*, Vol. 18(1), 1–6. https://doi.org/10.52939/ijg.v18i1.2095.

[15] Arrasyid, R., Ihsan, H., Darsiharjo, Ruhimat, M. and Pratama, A., (2023). Suitability Evaluation of Land Use/ Land Cover (LULC) Towards Landslide Prone Areas in Structural and Volcano Landform. *International Journal of Geoinformatics*, Vol. 19(6), 61–75. https://doi.org/10.52939/ijg.v19i6.2697.

[16] Xu, Y., Ouyang, C., Xu, Q., Wang, D., Zhao, B. and Luo, Y., (2024). CAS Landslide Dataset: A Large-Scale and Multisensor Dataset for Deep Learning-Based Landslide Detection. *Scientific Data*, Vol. 11(1), 1–11. https://doi.org/10.1038/s41597-023-02847-z.

[17] Mezaal, M. R. and Pradhan, B., (2018). An Improved Algorithm for Identifying Shallow and Deep-seated Landslides in Dense Tropical Forest from Airborne Laser Scanning Data. *Catena*, Vol. 167, 147–159. https://doi.org/10.1016/j.catena.2018.04.038.

[18] Shin, H. C., Roth, H. R., Gao, M., Lu, L., Xu, Z., Nogues, I., Yao, J., Mollura, D., and Summers, R. M. (2016). Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning. *IEEE Transactions on Medical Imaging*, Vol. 35(5), 1285–1298. https://doi.org/10.1109/TMI.2016.2528162.

[19] Bragagnolo, L., Rezende, L. R., da Silva, R. V. and Grzybowski, J. M. V., (2021). Convolutional Neural Networks Applied to Semantic Segmentation of Landslide Scars. *Catena*, Vol. 201. https://doi.org/10.1016/j.catena.2021.105189.

[20] Ghorbanzadeh, O., Blaschke, T., Gholamnia, K., Meena, S. R., Tiede, D. and Aryal, J., (2019). Evaluation of Different Machine Learning Methods and Deep-learning Convolutional Neural Networks for Landslide Detection. *Remote Sensing*, Vol. 11(2). https://doi.org/10.3390/rs11020196.

[21] Oak, O., Nazre, R., Naigaonkar, S., Sawant, S. and Vaidya, H., (2024). *A Comparative Analysis of CNN-based Deep Learning Models for Landslide Detection*. Available: http://arxiv.org/abs/2408.01692. [Accessed: May 8, 2024].

[22] Şener, A. and Ergen, B., (2024). LandslideSegNet: An Effective Deep Learning Network for Landslide Segmentation using Remote Sensing Imagery. *Earth Science Informatics*, Vol. 3963–3977. https://doi.org/10.1007/s12145-024-01434-z.

[23] Wei, R., Ye, C., Ge, Y. and Li, Y., (2022). An Attention-constrained Neural Network with Overall Cognition for Landslide Spatial Prediction. *Landslides*, Vol. 19(5), 1087–1099. https://doi.org/10.1007/s10346-021-01841-z.

[24] Wang, X., Wang, D., Sun, T., Dong, J., Xu, L., Li, W., Li, S., Ran, P., Ao, J., Zou, Y., Wang, J. and Zeng, X., (2023). Dual Path Attention Network (DPANet) for Intelligent Identification of Wenchuan Landslides. *Remote Sensing*, Vol. 15(21). https://doi.org/10.3390/rs15215213.

[25] Amankwah, S. O. Y., Wang, G., Gnyawali, K., Hagan, D. F. T., Sarfo, I., Zhen, D., Nooni, I. K., Ullah, W. and Duan, Z. (2022). Landslide Detection from Bitemporal Satellite Imagery using Attention-based Deep Neural Networks. *Landslides*, Vol. 19(10), 2459–2471. https://doi.org/10.1007/s10346-022-01915-6.

[26] Sun, Y., Bi, F., Gao, Y., Chen, L. and Feng, S., (2022). A Multi-Attention UNet for Semantic Segmentation in Remote Sensing Images. *Symmetry*, Vol. 14(5), 1–19. https://doi.org/10.3390/sym14050906.

[27] Thammaboribal, P., Triapthti, N., and Lipiloet, S. (2025). Using of Analytical Hierarchy Process (AHP) in Disaster Management: A Review of Flooding and Landslide Susceptibility Mapping. *International Journal of Geoinformatics*, Vol. 21(4), 177–196. https://doi.org/10.52939/ijg.v21i4.4091.

[28] Fadlin, F. and Rahmalia, T., (2014). Study of the Relationship Level of Alteration on the Potential of Landslides Based on Analysis Petrography and X-ray Diffraction Along the Arjosari-Tegalombo Road, Pacitan District. *Prosiding Seminar Nasional ReTII Ke-9, 2014*, *19*.

[29] Ji, S., Yu, D., Shen, C., Li, W. and Xu, Q., (2020). Landslide Detection from an Open Satellite Imagery and Digital Elevation Model Dataset using Attention Boosted Convolutional Neural Networks. *Landslides*, Vol. 17(6), 1337–1352. https://doi.org/10.1007/s10346-020-01353-2.

[30] Alqaraawi, A., Schuessler, M., Weiß, P., Costanza, E. and Berthouze, N., (2020). Evaluating Saliency Map Explanations for Convolutional Neural Networks. *International Conference on Intelligent User Interfaces, Proceedings IUI*, 275–285. https://doi.org/10.1145/3377325.3377519.

[31] Corominas, J., van Westen, C., Frattini, P., Cascini, L., Malet, J. P., Fotopoulou, S., Catani, F., Van Den Eeckhaut, M., Mavrouli, O., Agliardi, F., Pitilakis, K., Winter, M. G., Pastor, M., Ferlisi, S., Tofani, V., Hervás, J. and Smith, J. T., (2014). Recommendations for the Quantitative Analysis of Landslide Risk. *Bulletin of Engineering Geology and the Environment*, Vol. 73(2), 209–263. https://doi.org/10.1007/s10064-013-0538-8.

[32] Olaf Ronneberger, Philipp Fischer, and T. B. (2015). Medical Image Computing and Computer-Assisted Intervention - MICCAI 2015. *18th International Conference Munich, Germany, October 5-9, 2015 Proceedings, Part III*. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* Vol. 9351. https://doi.org/10.1007/978-3-319-24574-4.

[33] Lu, W., Hu, Y., Zhang, Z. and Cao, W., (2023). A Dual-encoder U-Net for Landslide Detection using Sentinel-2 and DEM Data. *Landslides*, Vol. 20(9), 1975–1987. https://doi.org/10.1007/s10346-023-02089-5.

[34] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł. and Polosukhin, I., (2017). Attention is All You Need. *Advances in Neural Information Processing Systems*, https://doi.org/10.48550/arXiv.1706.03762.

[35] Wang, J., Chen, G., Jaboyedoff, M., Derron, M. H., Li Fei, Li, H. and Luo, X., (2023). Loess Landslides Detection via a Partially Supervised Learning and Improved Mask-RCNN with Multi-source Remote Sensing Data. *Catena*, Vol. 231. https://doi.org/10.1016/j.catena.2023.107371.

[36] Taalab, K., Cheng, T. and Zhang, Y., (2018). Mapping Landslide Susceptibility and Types using Random Forest. *Big Earth Data*, Vol. 2(2), 159–178. https://doi.org/10.1080/20964471.2018.1472392.

[37] Salehi, S. S. M., Erdogmus, D. and Gholipour, A., (2017). Tversky Loss Function for Image Segmentation using 3D Fully Convolutional Deep Networks. *International Workshop on Machine Learning in Medical Imaging*. Cham: Springer International Publishing, 379-387.

[38] Șandric, I., Chițu, Z., Ilinca, V. and Irimia, R., (2023). Using High Resolution UAV Imagery and AI to Automatically Detect and Map Landslide Features. *Remote Sensing of Environment*, Vol. *January*. https://doi.org/10.1007/s10346-024-02295-9.

[39] Terven, J., Cordova-Esparza, D. M., Ramirez-Pedraza, A., Chavez-Urbiola, E. A. and Romero-Gonzalez, J. A., (2023). *Loss Functions and Metrics in Deep Learning*. [Online]. Available: http://arxiv.org/abs/2307.02694. [Accessed: Apr 15, 2024].

[40] Simonyan, K., Vedaldi, A. and Zisserman, A., (2014). Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps. *2nd International Conference on Learning Representations, ICLR 2014 - Workshop Track Proceedings*, 1–8.

[41] Sreelakshmi, S. and Vinod Chandra, S. S., (2024). Visual Saliency-based Landslide Identification using Super-resolution Remote Sensing Data. *Results in Engineering*, Vol. 21. https://doi.org/10.1016/j.rineng.2023.101656.

[42] Safonova, A., Ghazaryan, G., Stiller, S., Main-Knorn, M., Nendel, C. and Ryo, M., (2023). Ten Deep Learning Techniques to Address Small data Problems with Remote Sensing. *International Journal of Applied Earth Observation and Geoinformation*, Vol. 125. https://doi.org/10.1016/j.jag.2023.103569.

[43] Woo, S., Park, J., Lee, J. Y. and Kweon, I. S., (2018). CBAM: Convolutional Block Attention Module. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Vol. 11211 LNCS, 3–19. https://doi.org/10.1007/978-3-030-01234-2_1.

[44] Brauwers, G. and Frasincar, F., (2023). A General Survey on Attention Mechanisms in Deep Learning. *IEEE Transactions on Knowledge and Data Engineering*, Vol. 35(4),

3279–3298. https://doi.org/10.1109/TKDE.2021.3126456.

[45] Oktay, O., Schlemper, J., Folgoc, L. Le, Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N. Y., Kainz, B., Glocker, B. and Rueckert, D., (2018). Attention U-Net: Learning Where to Look for the Pancreas. *1st Conference on Medical Imaging with Deep Learning*, *Midl*. https://doi.org/arXiv preprint arXiv:1804.03999.

[46] Wang, X., Wang, D., Sun, T., Dong, J., Xu, L., Li, W., Li, S., Ran, P., Ao, J., Zou, Y., Wang, J. and Zeng, X., (2023). Dual Path Attention Network (DPANet) for Intelligent Identification of Wenchuan Landslides. *Remote Sensing*, Vol. 15(21). https://doi.org/10.3390/rs15215213.

[47] Reyes, A. A., Paheding, S., Rajaneesh, A., Sajinkumar, K. S. and Oommen, T., (2023). C-PLES: Contextual Progressive Layer Expansion with Self-attention for Multi-class Landslide Segmentation on Mars using Multimodal Satellite Imagery. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, Vol. *2023-June*, 354–364. https://doi.org/10.1109/CVPRW59228.2023.00041.

[48] Nagendra, S., Shen, C. and Kifer, D., (2023*). Estimating Uncertainty in Landslide Segmentation Models*. arXiv preprint arXiv:2311.11138.

[49] Phakdimek, S., Komori, D. and Chaithong, T., (2023). Combination of Optical Images and SAR Images for Detecting Landslide Scars, using a Classification and Regression Tree. *International Journal of Remote Sensing*, Vol. 44(11), 3572–3606. https://doi.org/10.1080/01431161.2023.2224096.

[50] Ma, Z., Mei, G. and Piccialli, F., (2021). Machine Learning for Landslides Prevention: A Survey. *Neural Computing and Applications*, Vol. 33(17), 10881–10907. https://doi.org/10.1007/s00521-020-05529-8.

[51] Hu, J., (2018). Squeeze and Excitation Networks CVPR-2018. *Cvpr*, 7132–7141. [Online]. Available: http://openaccess.thecvf.com/content_cvpr_2018/html/Hu_Squeeze-and-Excitation_Networks_CVPR_2018_paper.html. [Accessed: May 10, 2024]

[52] Yu, C., Wang, J., Peng, C., Gao, C., Yu, G. and Sang, N., (2018). BiSeNet: Bilateral Segmentation Network for Real-time Semantic Segmentation. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Vol. 11217(LNCS), 334–349. https://doi.org/10.1007/978-3-030-01261-8_20

[53] Ghorbanzadeh, O. and Blaschke, T., (2019). Optimizing Sample Patches Selection of CNN to Improve the MIOU on Landslide Detection. *GISTAM 2019 - Proceedings of the 5th International Conference on Geographical Information Systems Theory, Applications and Management*, Vol. (*May)*, 33–40. https://doi.org/10.5220/0007675300330040.

[54] Hamwood, J., Alonso-Caneiro, D., Read, S. A., Vincent, S. J. and Collins, M. J., (2018). Effect of Patch Size and Network Architecture on a Convolutional Neural Network Approach for Automatic Segmentation of OCT Retinal Layers. *Biomedical Optics Express*, Vol. 9(7). https://doi.org/10.1364/boe.9.003049.

[55] Reichenbach, P., Rossi, M., Malamud, B. D., Mihir, M. and Guzzetti, F., (2018). A Review of Statistically-Based Landslide Susceptibility Models. *Earth-Science Reviews*, Vol. 180, 60–91. https://doi.org/10.1016/j.earscirev.2018.03.001.

[56] Kikuchi, T., Sakita, K., Nishiyama, S. and Takahashi, K., (2023). Landslide Susceptibility Mapping using Automatically Constructed CNN Architectures with Pre-slide Topographic DEM of Deep-seated Catastrophic Landslides Caused by Typhoon Talas. *Natural Hazards*, Vol. 117(1), 339–364. https://doi.org/10.1007/s11069-023-05862-w.

[57] Chattopadhay, A., Sarkar, A., Howlader, P. and Balasubramanian, V. N., (2018). Grad-CAM++: Generalized Gradient-based Visual explanations for deep convolutional networks. *Proceedings - 2018 IEEE Winter Conference on Applications of Computer Vision, WACV 2018*, Vol. 2018-January; 839–847. https://doi.org/10.1109/WACV.2018.00097.

[58] Pradhan, B., Al-Najjar, H. A. H., Sameen, M. I., Mezaal, M. R. and Alamri, A. M., (2020). Landslide Detection Using a Saliency Feature Enhancement Technique from LiDAR-Derived DEM and Orthophotos. *IEEE Access*, Vol. 8, 121942–121954. https://doi.org/10.1109/ACCESS.2020.3006914.

[59] Paudel, U., Oguchi, T. and Hayakawa, Y., (2016). Multi-Resolution Landslide Susceptibility Analysis Using a DEM and Random Forest. *International Journal of Geosciences*, Vol. 07(05). 726–743. https://doi.org/10.4236/ijg.2016.75056.