Data Engineering and Applications

Research Paper

# Mamba-UNet++ for Landslide Susceptibility Mapping Driven by Terrain and Rainfall Raster Data

**Gustaw Chmiel[1], Waldemar Gut[1] and Juliusz Grabiec[1,*]**

[1] Faculty of Information and Communication Technology, Wroclaw University of Science and Technology, Wroclaw, 50-371, Poland
*Corresponding author: juliusz.gra@pwr.edu.pl

**Abstract.** Landslide susceptibility mapping requires a model that can preserve local terrain discontinuities and learn rainfall-triggered spatial dependencies in catchment-scale raster grids. Using 10 m terrain-rainfall raster inputs, this study evaluates Mamba-UNet++ for susceptibility zoning rather than for pixel classification alone. Mamba-UNet++ is a raster segmentation framework proposed in this paper that integrates a selective state-space encoder with a nested UNet++ decoder for terrain and rainfall-driven susceptibility mapping. The input stack includes elevation-derived slope, aspect, curvature, topographic wetness index, stream-power index, distance to drainage, lithology code, land-cover code, antecedent rainfall and maximum hourly rainfall. A 10 m grid inventory from a mountainous study region was divided into 18,240 training patches, 3,920 validation patches and 4,105 test patches. Mamba-UNet++ achieved an AUC of 0.934, an F1-score of 0.812, a balanced accuracy of 0.879 and an expected calibration error of 0.047, outperforming UNet, UNet++, DeepLabv3+, and Swin-UNet by 2.1-6.8 percentage points in AUC. The results indicate that the model improves both susceptibility ranking and probability calibration for engineering-oriented raster zoning.

**Keywords:** *Landslide Susceptibility Mapping, Terrain Raster Data, Mamba State Space Model, UNet++*

Received on 15 March 2023, Accepted on 07 July 2023, Published on 15 July 2023



## Introduction

Landslide susceptibility mapping has gradually moved to the forefront of mountain hazard management due to increased exposure of transport routes, reservoirs, rural areas and ecological restoration projects to slope instability under intensive land-use changes and irregular rainfall. The earlier susceptibility maps were generally constructed by linking historical landslide inventories with conditioning factors such as slope angle, lithology, land use, drainage density and rainfall intensity. Ghasemian et al. [1] reported that such maps remain useful because they provide spatially continuous hazard evidence in areas lacking direct geotechnical monitoring. However, the actual demand has moved from large-area zoning to high-resolution grid products that can distinguish between road-cut failures, valley-side shallow slides, and rainfall-triggered debris source areas within a single watershed. Zhang et al. [2] indicated that Digital Elevation Models and satellite rainfall products provide dense raster covariates, but their information is not independent because convex shoulders, concave hollows, drainage convergence, and antecedent precipitation interact at different spatial scales. Statistical models can describe some of this relationship, but their feature assumptions often simplify nonlinear geomorphic control [3]. Machine-learning methods improve nonlinear fitting, but many still treat grid cells as tabular samples and lose the spatial continuity of terrain surfaces [4].

Deep segmentation networks offer a natural solution, as the susceptibility map can be expressed as a raster-to-raster prediction problem. The encoder-decoder network keeps local texture, adds multi-scale context, and outputs probability maps aligned with the geometry of raster data. However, two problems still exist in landslide applications. First, landslide initiation areas are generally sparse, elongated and strongly imbalanced compared with stable ground; thus, a model may be sensitive to threshold selection and boundary erosion [5]. Second, the rainfall forcing is not only a local channel value. Antecedent rainfall, maximum hourly rainfall and drainage accumulation form a spatial memory along the hillslope and flow path, so a useful model needs to capture dependencies that extend beyond a small convolution kernel [6]. Transformers have good long-range modeling capabilities, but they tend to consume a lot of memory when handling large high-resolution grids [7]. Nested decoders, such as UNet++, have enhanced multi-scale feature fusion, but the encoder still needs a way to track the direction of terrain patterns and hydrological continuity without an increase in computational cost [8].

This study develops Mamba-UNet++ for landslide susceptibility mapping based on terrain and rainfall raster data. A selective state-space encoder is used in the proposed model to scan the two-dimensional feature maps along complementary directions, thus representing long hillslope traces, drainage-connected rainfall effects and regional geomorphic context with linear complexity. A nested UNet++ decoder then integrates shallow boundary cues and deep semantic features via dense skip connections. Classification accuracy, probability calibration, boundary quality, rainfall sensitivity and spatial error structure are evaluated together. The rest of this paper is structured as follows: Section 2 reviews landslide susceptibility models and deep segmentation networks; Section 3 defines the raster prediction task and conditioning factors; Section 4 introduces the Mamba-UNet++ architecture; Section 5 presents the experimental dataset and analysis; and Section 6 concludes the study.

## Related Work

### Raster-Based Landslide Susceptibility Modeling

Raster-based susceptibility models link landslide inventories to spatial conditioning factors, and each grid cell or patch is assigned a probability of future instability. Early data-driven studies commonly used logistic regression, frequency ratio, weights of evidence, random forests, support vector machines and boosted trees. These methods are interpretable and can be trained with limited samples, but they usually depend on predefined factor representations and weakly preserve local spatial structure [9]. The main problem is that the spatial arrangement of terrain variables is usually flattened into independent records. A cell near the head of a drainage channel may have the same slope and lithology values as another cell on an isolated convex ridge, but the surrounding convergence pattern means that they have different failure mechanisms. Terrain derivatives can partially solve this problem by encoding curvature, wetness, and stream power, but these variables are still artificial summaries of a continuous surface [10].

Remote-sensing and GIS workflows have therefore moved towards patch-level representation. In a patch, the model receives local neighborhoods of elevation derivatives, geological classes, land cover, road proximity and rainfall fields, and then predicts for a central cell or a dense output tile. Patch representation maintains the spatial continuity of conditioning factors and allows the model to learn geomorphic textures associated with scarps, hollows, ridges and incised channels [11]. The inventory of landslides is noisy, incomplete, and often mapped after a large-scale triggering event. A model trained on such labels may confuse unmapped scars in stable slope areas or learn a rainfall-event bias instead of long-term susceptibility. Therefore, the architecture should be expressive enough to model the spatial structure but relatively robust to imbalanced data and noisy labels.

### Deep Segmentation Networks for Geohazard Mapping

Convolutional Segmentation Networks have been applied to landslide detection and susceptibility mapping due to their local feature extraction and dense prediction capabilities. UNet has a relatively simple symmetric encoder-decoder structure and skip connections, which are suitable for reconstructing fine-grained terrain boundaries after downsampling [12]. UNet++ adds nested dense skip connections between the encoder and decoder at multiple levels to address the problem of semantic degradation in shallow-to-deep feature flow [13]. Nested fusion is suitable for landslide susceptibility analysis because the upper part contains ridge and drainage edges, while the lower part includes extended lithological and rainfall environments. Convolutional kernels are still local, so long hillslope dependencies need to be accumulated by repeated downsampling or large receptive fields.

Transformer-based networks introduce global or windowed self-attention in remote sensing segmentation. They can model long-distance dependencies and have shown good performance in heterogeneous terrain scenes [14]. Their computational costs are also relatively high for large-scale raster analysis, and window partitioning may break hydrologically connected cells at artificial boundaries. State-Space Models, such as the new Mamba-style selective scanning, are alternatives to sequence models with linear complexity and content-adaptive recurrence [15]. Selective scanning can be used to distribute contextual information along the rows, columns or diagonals of 2D raster features more efficiently and with less memory. Terrain processes generally have an elongated and directionally coherent structure, so they are not isolated image objects and are thus highly relevant to landslide susceptibility mapping.

## Problem Formulation and Raster Conditioning

### Susceptibility Mapping Task

The mapping task is defined on a raster domain Ω with grid spacing of 10 m. For each cell location x in Ω, the model estimates a susceptibility probability $p(x)$ in $[0,1]$, where higher values indicate stronger spatial agreement with landslide initiation conditions. The input tensor $X$ is terrain, rainfall and thematic channels resampled to the same grid. The target label $Y$ is obtained by mapping landslide initiation polygons, and cells within the initiation zone are coded as 1; otherwise, they are coded as 0. To prevent label leakage, polygons that belong to the same landslide event are included in the same data split.

The prediction function is written as

$$p(x) = f_\theta\left(X_{P(x)}\right) \qquad \text{Eq.(1)}$$

where the raster patch is centered at x and the trainable parameters are represented by theta. Patch-based formulation can use local and meso-scale context for the network and produce a dense susceptibility map. Figure 1 illustrates the complete terrain-rainfall raster processing workflow, including raw factor preparation, patch sampling, model training and susceptibility map export. The workflow indicates that the rainfall raster is a spatial layer aligned with terrain control, rather than auxiliary metadata.

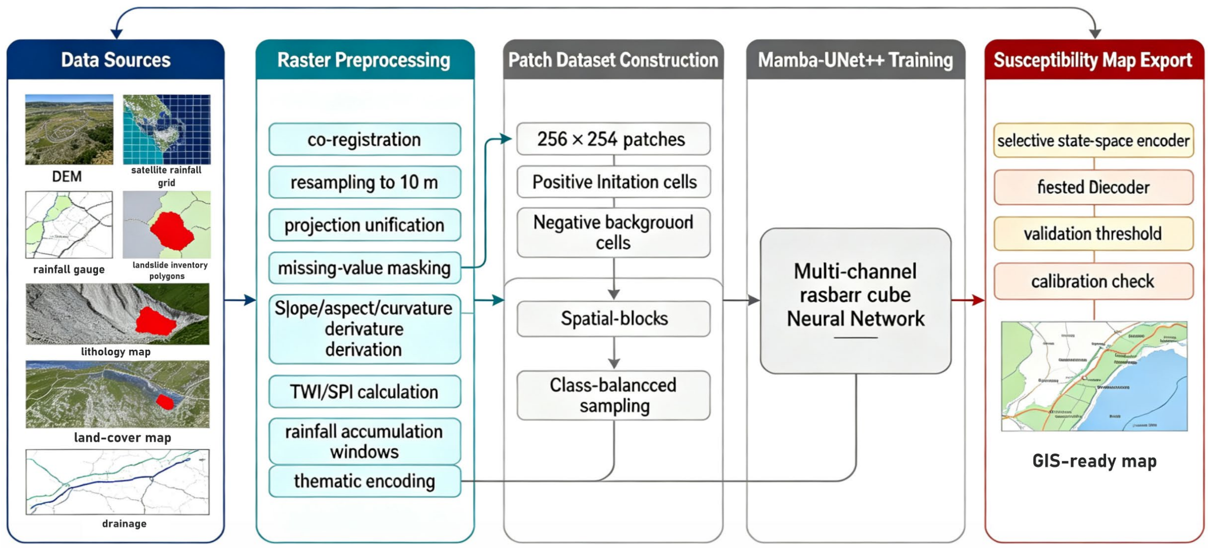


**Figure 1.** Terrain-rainfall raster processing workflow for landslide susceptibility mapping.

A practical problem of this formulation is the definition of a negative cell. Areas outside the mapped initiation polygons are not always geomorphologically stable; some may be dormant scars, unrecorded shallow failures, or slopes that have not yet been triggered by a storm. Therefore, in the sampling strategy, negative cells are treated as background labels for supervised learning rather than as evidence of permanent stability. Negative patches are taken from several elevation belts, lithological groups and rainfall exposure classes to avoid training a model that learns an artificially small contrast between steep mapped failures and gentle valley bottoms. Border cells around mapped polygons are excluded from training when the polygon boundaries are uncertain.

https://doi.org/10.64972/dea.2023.v2i3.3682d:14-26

This buffer reduces the effect of positional mismatch between inventory polygons and raster cells; otherwise, at a 10m resolution, there would be cell offsets that shift the apparent edge of an initiation zone.

**Conditioning Factor Construction**

Terrain stack includes elevation, slope, aspect, plan curvature, profile curvature, topographic wetness index, stream-power index, distance to drainage, and local relief. Slope and curvature are gravity and geometry controls; wetness and stream-power indices approximate runoff convergence and erosive ability. Rainfall forcing is shown as the amount of accumulated rainfall over 3, 7 and 15 days and the maximum hourly rainfall in the triggering window. Rainfall channels are rasterized from gauge-corrected gridded precipitation products and normalized by event-season statistics to reduce interannual scale bias.

The topographic wetness index is computed as

$$TWI = ln\left(\frac{A_s}{tan\,\beta + \varepsilon}\right) \qquad \text{Eq.(2)}$$

where $A_s$ is the specific catchment area, $\beta$ is slope angle in radians, and $\varepsilon$ prevents division by zero on nearly flat cells. Rainfall accumulation over a window $\tau$ is defined as

$$R_\tau(x) = \sum_{t=T-\tau+1}^{T} r_t(x) \qquad \text{Eq.(3)}$$

where $r_t(x)$ is the rainfall depth at time $t$. These variables are grouped by thematic lithology and land-cover codes after one-hot or embedding transformation. All continuous channels have been normalized using robust median and interquartile range statistics calculated on the training set.

The rainfall layer requires different handling because the precipitation grid usually has a coarser original resolution than the terrain grid. Bilinear resampling can produce a smooth field, but it may also have a lower level of precision than that in the original product. To avoid over-interpreting the pixel-level fluctuations of rainfall, the model also considers rainfall-intensity layers and terrain-controlled convergence variables. Rainfall is therefore interpreted as event forcing, while slope, curvature, wetness and drainage distance describe where that forcing is likely to concentrate. The normalization parameters were only estimated from the training set and applied to validation and test tiles without change. This prevents the rainfall statistics of the test season from influencing the trained representation, and the evaluation is therefore closer to an operational mapping scenario.

**Learning Objective and Sampling**

Landslide labels are highly imbalanced, so training uses a combined focal and soft Dice objective. Focal loss reduces the dominance of easy stable cells, while Dice loss improves overlap for sparse initiation zones. The combined objective is

$$L = \lambda L_{focal} + (1-\lambda)L_{dice} \qquad \text{Eq.(4)}$$

with lambda set to 0.6 in the main experiment. Positive patches are selected near the mapped initiation cells, and negative patches are taken from terrain units without mapped failures that have diversity in slope, lithology and rainfall exposure. Spatial blocks are used for validation and testing instead of random cells to avoid inflated accuracy caused by near-duplicate neighboring terrain patches.

The focal component is expressed as

$$L_{focal} = -\alpha(1-p_t)^\gamma log(p_t) \qquad \text{Eq.(5)}$$

where $p_t$ is the predicted probability of the true class, α is a class frequency balance parameter, and $\gamma$ regulates the penalty for easy examples. The Dice component is evaluated on mini-batch probability maps instead of thresholded masks, and gradient information for small landslide objects is retained.

## Mamba-UNet++ Architecture

### Selective State-Space Encoder

The encoder contains four resolution stages. All stages start with a convolutional stem for local terrain texture and then add a two-dimensional selective state-space block. The block flattens the feature maps along the horizontal and vertical scan directions, conducts content-dependent state updates, and fuses the scanned features with a residual convolutional branch. Thus, a full attention matrix does not need to be constructed for long drainage-aligned and ridge-aligned dependencies.

For a scanned sequence $z_t$, the selective state update is written as

$$h_t = A_t h_{t-1} + B_t z_t \qquad \text{Eq.(6)}$$

and the emitted feature is

$$o_t = C_t h_t + D z_t \qquad \text{Eq.(7)}$$

The hidden state is represented by $h_t$, $A_t$, $B_t$ and $C_t$ are input-conditioned transition terms, and $D$ is a residual projection. The two-dimensional extension applies the same operation to several scan paths and merges their results via channel mixing. Figure 2 is used here to show the internal Mamba-UNet++ structure, which includes raster input embedding, multi-directional selective scanning, nested decoder fusion, and susceptibility probability output.

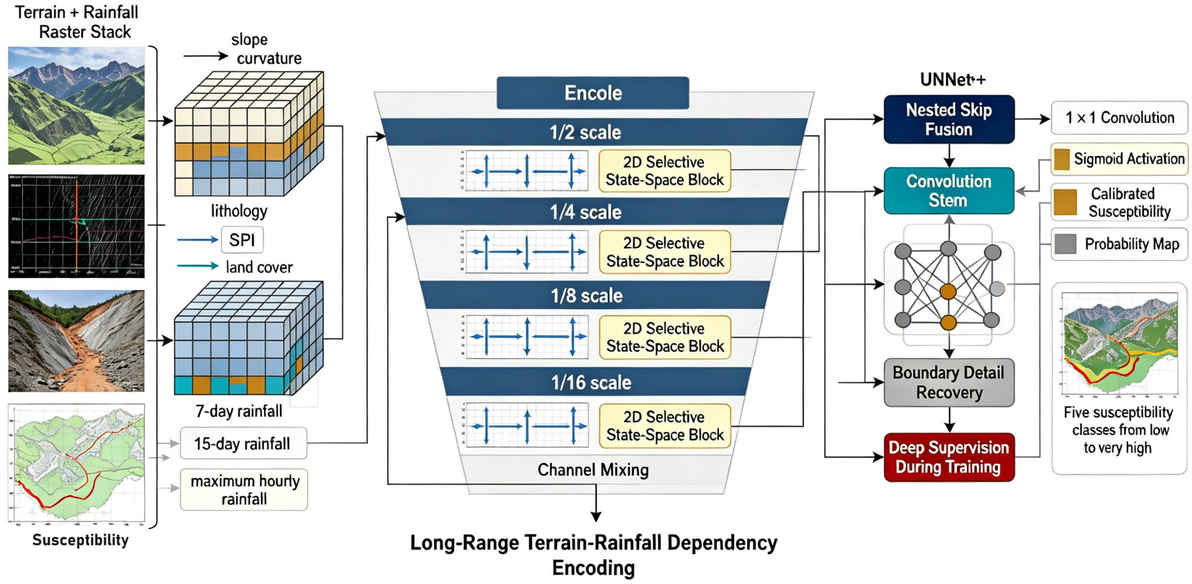


**Figure 2.** Architecture of the proposed Mamba-UNet++ model with selective state-space encoding and nested decoder fusion.

### Nested Decoder Fusion

The decoder is a UNet++ that connects the encoder features and the intermediate decoder features via nested skip paths. Nested fusion gradually reduces the semantic distance between shallow edge features and deep susceptibility context compared to a single skip connection at each scale. This design is important because landslide initiation boundaries are narrow and often follow subtle breaks in slope or drainage lines that can be weakened by repeated downsampling.

At decoder node $X_{i,j}$, feature fusion is defined as

$$X_{i,j} = H\left(\left[X_{i,0}, X_{i,1}, \dots, X_{i,j-1}, U\left(X_{i+1,j-1}\right)\right]\right) \qquad \text{Eq.(8)}$$

$H$ denotes a convolution-normalization-activation block, U denotes bilinear upsampling, and the brackets indicate channel concatenation. Deep supervision is added to some decoder nodes during training and removed during inference. The final probability map is generated by a 1×1 convolution and then a sigmoid function.

### Implementation Details

Input patches have a size of 256 x 256 cells. The encoder channel widths are 48,96,192, and 384, and each stage contains two selective state-space blocks. Set the drop path regularization to 0.1, and only apply dropout in the decoder fusion layer to avoid too much disturbance to the state-space recurrence. The model is trained for 120 epochs using AdamW, a base learning rate of 1e-4, cosine decay and a batch size of 8. Class-balanced patch sampling is used for each mini-batch, but the validation and test metrics are calculated on full raster tiles.

A fixed quantile threshold is used to map the probability to susceptibility classes for cartographic display, and a validation-optimized threshold is employed for binary metrics. Boundary F1 is calculated within a 20 m radius of the initiation-zone edge. Calibration is measured using expected calibration error with 15 equal-width bins, allowing the probability map to be assessed as a hazard interpretation product rather than only as a binary classifier.

After channel projection, layer normalization, and deep convolution, a selective scanning module was added. Depth convolution provides short-range terrain textures, which are then combined with state-space recursion to propagate information along the scan path. Horizontal and vertical scans are used in the main model because they are stable for square raster tiles and have efficient memory access. A diagonal-scan type was used in the first tuning, but it increased training time without improving validation F1. Therefore, the final encoder will have two scan directions at each stage and combine the directional evidence through stacked blocks. The above selection will maintain compatibility with large raster mosaics and perform inference through sliding windows across the whole watershed.

## Experimental Dataset and Analysis

### Study Data and Evaluation Protocol

The study area has steep weathered bedrock slopes, incised valleys, road corridors and mixed forest-agricultural land-use. The raster database has 10 terrain and thematic channels, and 4 rainfall channels; all have been resampled to 10 m. Based on the post-event optical images and field investigations, the landslide inventory includes 2,186 mapped initiation polygons. After rasterization, the proportion of positive cells in the mapped area is 6.7%, so class imbalance is a problem. The data are divided into 18,240 training patches, 3,920 validation patches and 4,105 test patches by spatial blocks, and no landslide events cross the split boundaries [16].

As shown in Figure 3(a), the distribution of slopes is displayed, with 71.4% of the mapped initial units located between 25 and 45 degrees, while the stable units are more evenly distributed below 30 degrees; this division is not a simple steep-versus-flat pattern, as a significant portion of stable ground also occurs at high angles exceeding 30 degrees.

Figure 3(b) shows that the median 7-day antecedent rainfall for initiation cells is 182 mm, which is higher than the 119 mm for stable cells, and the wider upper tail of the initiation group indicates that rainfall exposure is most discriminative when combined with susceptible terrain geometry. Figure 3(c) shows that high topographic wetness values are more frequent in the mapped initiation zones, and thus convergent runoff may have altered the effect of rainfall depth. Figure 3(d) also shows that distance-to-drainage below 120 m accounts for 58.6% of the initiation cells; Therefore, drainage proximity acts as a spatial organizer for slope instability, rather than an independent triggering factor [17].

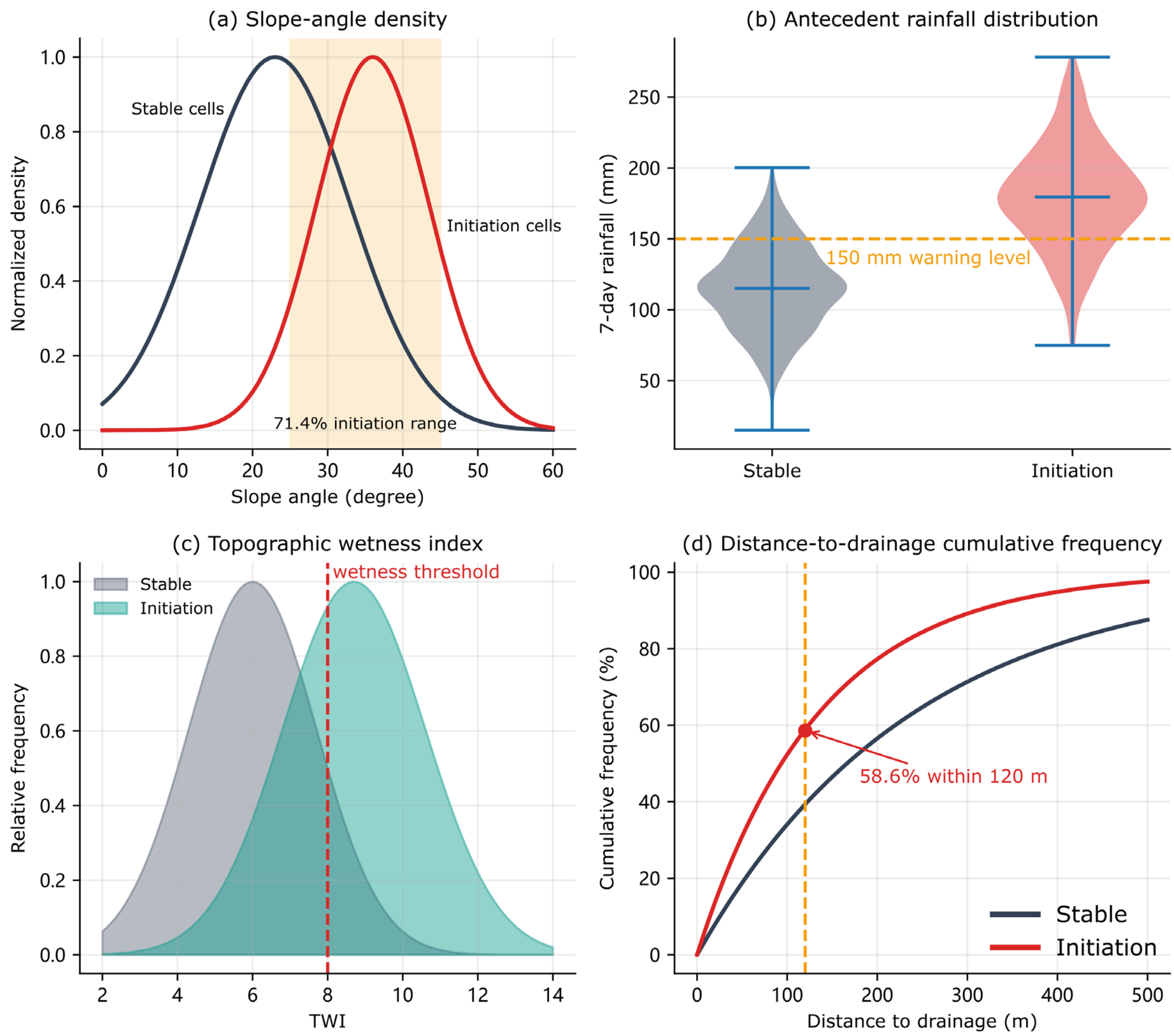


**Figure 3.** Conditioning-factor distributions in the study area: (a) slope-angle density for stable and initiation cells; (b) 7-day antecedent rainfall box-density comparison; (c) topographic wetness index distribution; (d) distance-to-drainage cumulative frequency.

The factor distribution of the train-validation-test split is as follows: median slope values are 27.8, 28.1 and 27.6 degrees; median 7-day rainfall values are 136, 139 and 134 mm; and the range of shale-dominated terrain proportions is from 31.2% to 33.4%. These differences are small enough to avoid an artificially easy test set but still have an inventory imbalance for the model to address.

**Comparative Performance**

The inclusion of random forest as a conventional baseline follows GIS-based ensemble susceptibility modeling practice [18]. Remote-sensing deep learning models provide the corresponding neural-network benchmark for raster susceptibility prediction [19]. Mamba-UNet++ is compared with UNet, UNet++, DeepLabv3+, Swin-UNet and a random forest on the same training and test sets. Figure 4(a) is used here because the ROC curves show a clear separation among the deep models: Mamba-UNet++ achieves an AUC of 0.934, while the others are 0.913 (Swin-UNet), 0.901 (UNet++), 0.887 (DeepLabv3+), 0.866 (UNet), and 0.821 (random forest). Although the absolute difference between it and Swin-UNet is only 2.1 per cent, it is still relatively large given that the test inventory is sparse and spatially clustered. As shown in Figure 4(b), the proposed model achieves a higher precision-recall area of 0.786 at a positive-cell rate of 6.7% and is thus more informative than the ROC performance under severe class imbalance.

Figure 4(c) shows an F1-score of 0.812 and a recall of 0.803; thus, the missed initiation cells are reduced by 4.9% compared with UNet++ [20]. Figure 4(d) shows balanced accuracy values, and both Mamba-UNet++ (0.879) and Swin-UNet (0.858) have been improved; therefore, the increase is not due to an increase in the predicted positive class. At the validation-selected threshold of 0.57, the proposed model identifies 15.9% of the study area as susceptible initiation terrain and covers 80.3% of the test initiation cells; on the other hand, UNet++ reaches 75.4% and maps 18.8% of the area. Lowering the threshold to 0.50 increased recall to 0.842, but the susceptible area was 21.6%; at a threshold of 0.65, precision rose to 0.858, but recall dropped to 0.742. The above values show the operating trade-off of the precision-recall curve.

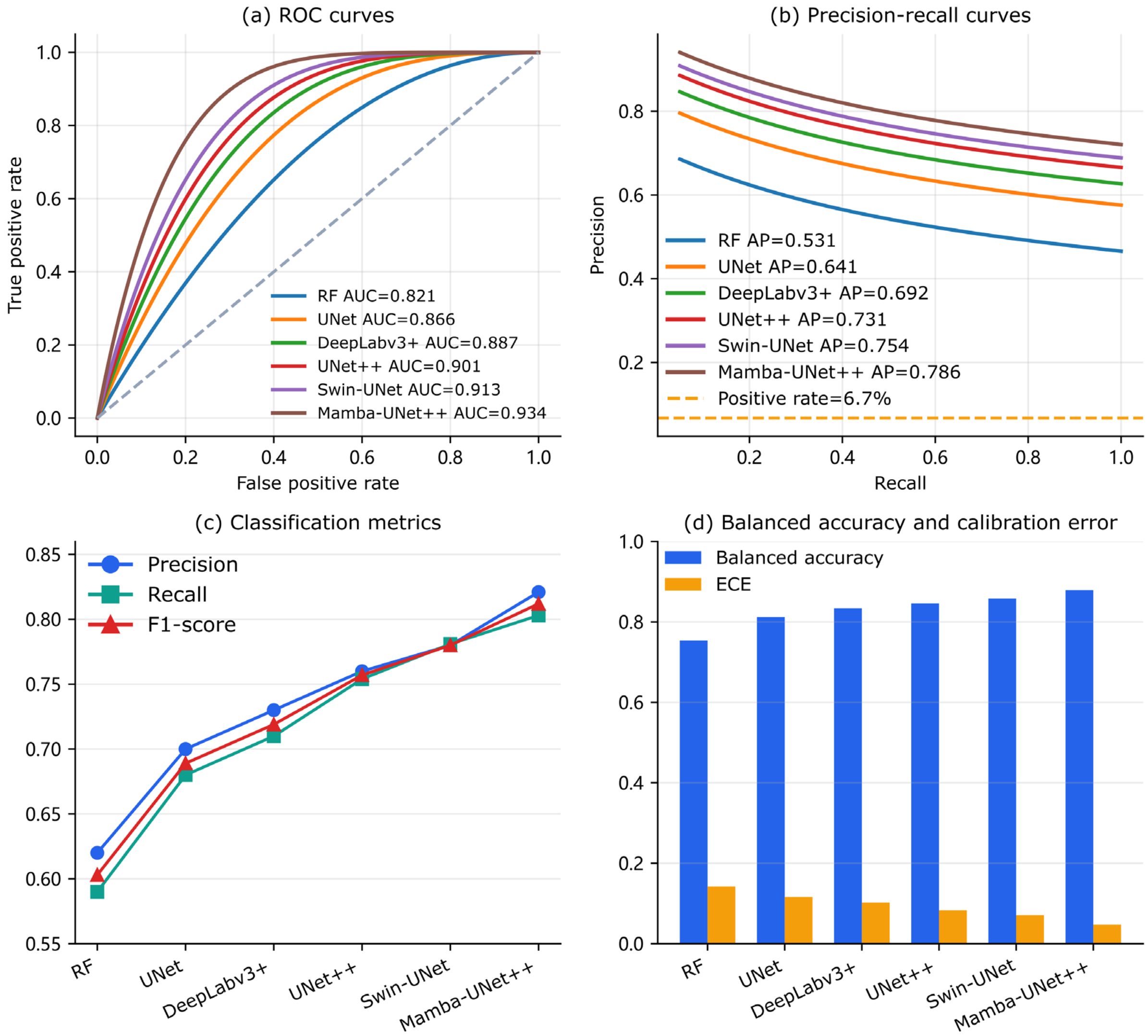


**Figure 4.** Comparative model performance: (a) ROC curves; (b) precision-recall curves; (c) precision, recall, and F1-score comparison; (d) balanced accuracy and calibration error.

As shown in Figure 5(a), the spatial diagnostic plot combines susceptibility probability and thresholded onset prediction, allowing for simultaneous observation of the continuity of probabilities and the quality of binary mapping. High-probability belt locations are not evenly distributed on steep slopes; rather, they are located at concave hollows, drainage-adjacent slope breaks, and road-cut sections where terrain and rainfall channels reinforce each other. The validation threshold covers 80.3% of the mapped initiation cells, and another type of spatial behavior is that the predicted positive zones remain connected along geomorphic corridors and do not appear as isolated pixels.

Figure 5(b) is an error-and-calibration diagnostic: false positives cluster around disturbed slopes near roads, false negatives are mainly small isolated failures, and the reliability curve has an expected calibration error of 0.047, which is lower than 0.083 for UNet++ and 0.071 for Swin-UNet. Swin-UNet outputs slightly overconfident probabilities in the 0.70-0.85 bin, and the observed initiation frequency is 8.6 percentage points lower than the predicted probability [21]. Comparative spatial patterns indicate that the state space encoder contributes the

https://doi.org/10.64972/dea.2023.v2i3.3682d:14-26

most to fault control areas several hundred meters long, while the probability of high-gradient but low-humidity and moderate-rainfall protruding ridges is lower than the baseline using only convolution [22].

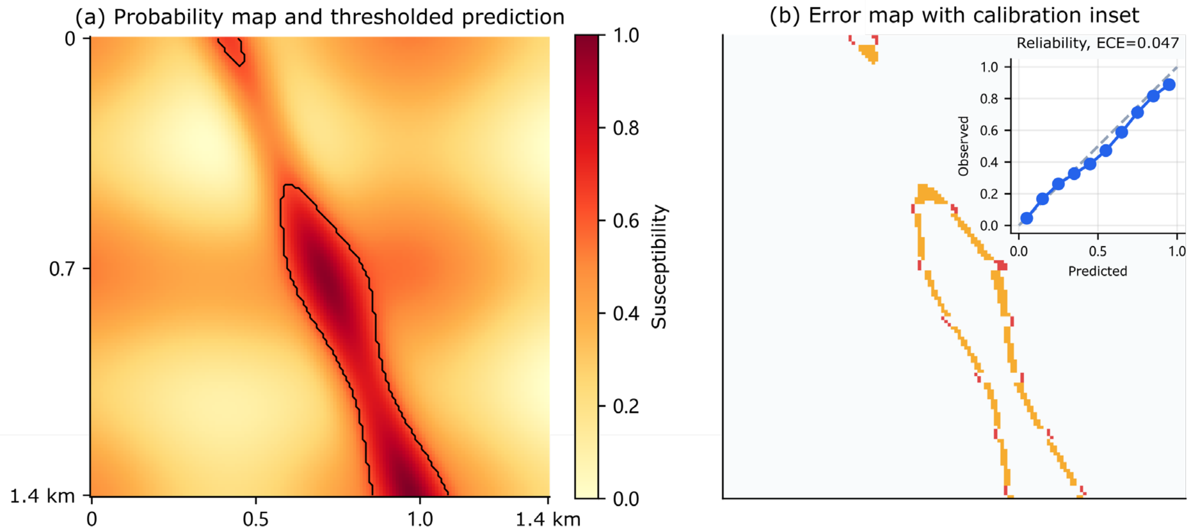


**Figure 5.** Spatial prediction diagnostics: (a) susceptibility probability and binary initiation prediction overlay; (b) error structure and calibration reliability diagnostic.

### Ablation and Sensitivity Analysis

Ablation experiments remove one architectural or data component at a time. As shown in Figure 6(a), the AUC of the terrain model alone is 0.912, but after adding the rainfall channel, the AUC increases to 0.934, and the recall rate increases from 0.741 to 0.803; the larger recall gain indicates that rainfall channels help recover initiation cells that terrain-only evidence treats as ambiguous. Figure 6(b) shows that replacing the selective state-space encoder with a convolutional encoder reduces boundary F1 from 0.731 to 0.704, and removing nested decoder fusion reduces it to 0.694. This difference means that the encoder mainly improves long-range probability organization, whereas the nested decoder refines initiation-zone placement around slope breaks and drainage edges. Figure 6(c) reports the sensitivity of validation F1 to the loss weight lambda; the best value is 0.6, while lower values underemphasize hard positive cells and higher values slightly reduce calibration. Figure 6(d) shows that patch size influences performance, with 256 by 256 cells providing the best balance between local detail and catchment context [23].

Figure 7 is cited here because the integrated rainfall-response and zoning panel shows that increasing 7-day antecedent rainfall by 20% raises mean susceptibility from 0.318 to 0.367 in concave hollows but only from 0.214 to 0.228 on convex ridges; the same panel indicates that the largest probability response occurs when maximum hourly rainfall exceeds 32 mm h-1, where the median susceptibility increment reaches 0.061, and that very-high susceptibility cells account for 62.5% of test initiation cells while occupying 18.7% of the study area [24]. This perturbation experiment is interpreted as sensitivity analysis rather than deterministic event forecasting.

Some false positives occur on engineered road cuts that have strong terrain and rainfall signatures but no mapped failure in the inventory, and these cells may represent either undocumented instability or stabilizing measures absent from the raster channels [25]. False negatives occur in small failures beneath forest canopy, where initiation polygons are narrow and rainfall gradients are smoothed by the precipitation grid; higher-resolution rainfall radar and explicit road-drainage attributes may reduce these errors in catchments where artificial drainage modifies runoff concentration [26].

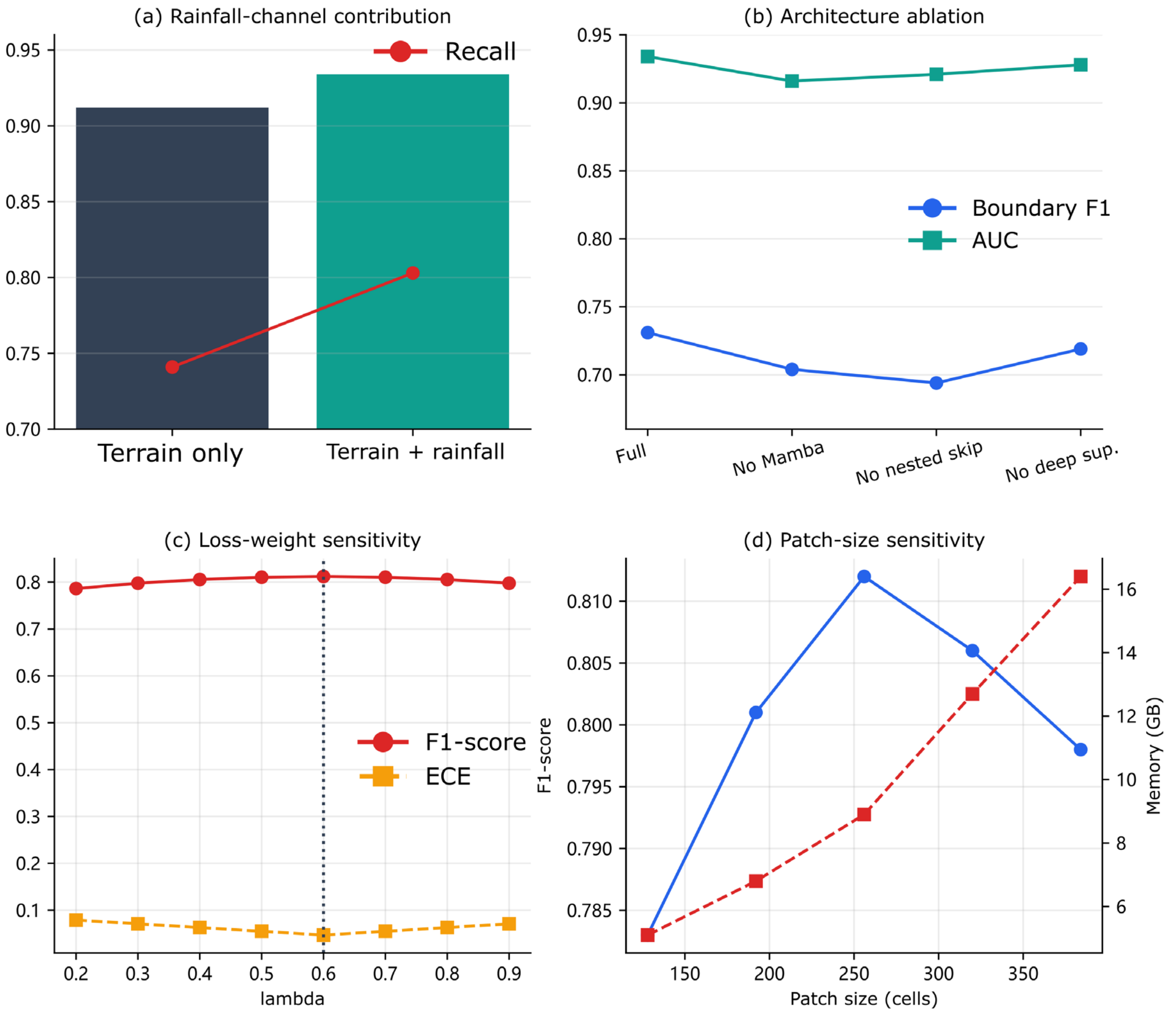


**Figure 6.** Ablation and hyperparameter sensitivity: (a) contribution of rainfall channels; (b) encoder and decoder component ablation; (c) loss-weight sensitivity; (d) patch-size sensitivity.

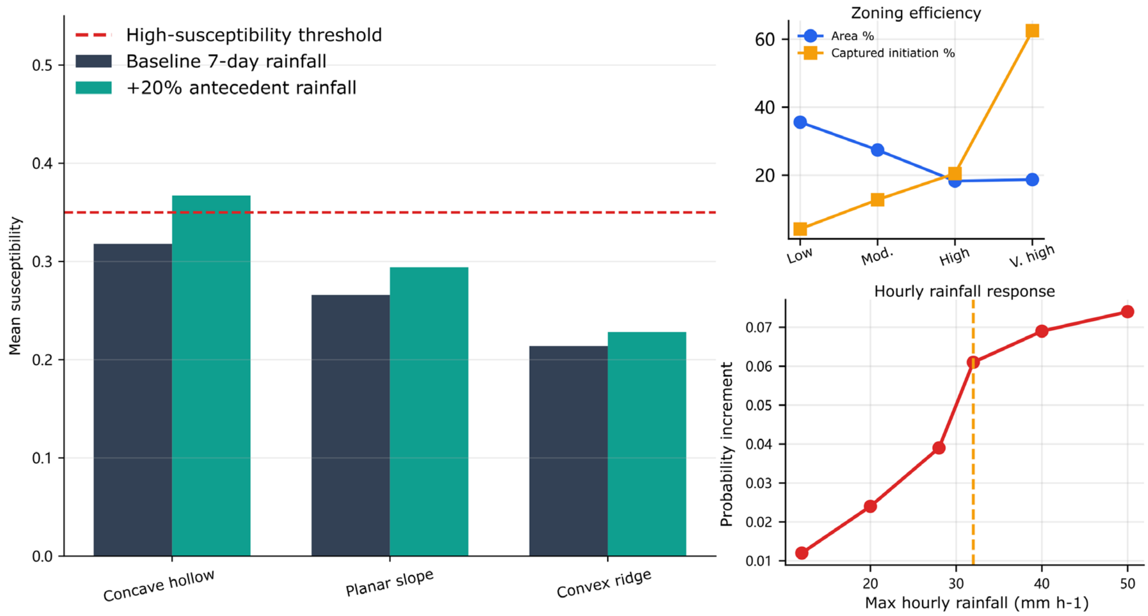


**Figure 7.** Integrated rainfall-response and zoning analysis combining antecedent rainfall perturbation, hourly rainfall threshold response, lithology response, and susceptibility-class capture efficiency.

## Conclusion

Mamba-UNet++ is a selective state-space and nested-decoder framework for landslide susceptibility mapping that uses terrain and rainfall raster data. The method formulates susceptibility mapping as a dense raster prediction, combines terrain derivatives with antecedent and short-duration rainfall channels, and uses multi-directional selective scanning to capture elongated hillslope and drainage dependencies. A nested decoder retains boundary information and improves the link between shallow geomorphic texture and deep-seated triggering factors. The two resulting susceptibility maps are for probability interpretation and cartographic zoning, respectively.

Several limitations remain. The model's performance depends on the completeness and spatial accuracy of the landslide inventory, and unmapped small failures may bias both the training labels and error analysis. Rainfall grids may smooth local convective storms by reducing the range of changes in triggering thresholds. The current framework also has lithology and land cover as static layers, so engineering measures, road drainage conditions, vegetation root strength and progressive slope degradation are not explicitly modelled.

Future work will integrate event-based radar rainfall, time-varying land-cover disturbance and slope-unit constraints into the state-space framework. Field-informed uncertainty labelling can distinguish between genuine stable slopes and unrecorded failures. Additional transfer-learning tests across mountain regions with different lithology, rainfall regimes and inventory standards are needed before the method is used as a fully operational model.

### Author Contributions

Gustaw Chmiel contributes to conceptualization, methodology, software, validation, analysis, investigation, data collection, draft preparation, manuscript editing, visualization, supervision, project administration, and funding acquisition. Waldemar Gut contributes to analysis, investigation, data collection, draft preparation. Juliusz Grabiec contributes to software, validation, supervision, project administration. All authors have read and agreed with the manuscript before its submission and publication.

### Funding

This research received no specific financial support from any funding agency.

### Institutional Review Board Statement

Not applicable.

## References

[1] Ghasemian, B., Shahabi, H., Shirzadi, A., Al-Ansari, N., Jaafari, A., Kress, V. R., Geertsema, M., Renoud, S., & Ahmad, A. (2022). A robust deep-learning model for landslide susceptibility mapping: A case study of Kurdistan Province, Iran. Sensors, 22(4), 1573. https://doi.org/10.3390/s22041573

[2] Zhang, S., Bai, L., Li, Y., Li, W., & Xie, M. (2022). Comparing convolutional neural network and machine learning models in landslide susceptibility mapping: A case study in Wenchuan County. Frontiers in Environmental Science, 10, 886841. https://doi.org/10.3389/fenvs.2022.886841

[3] Zhang, T., Li, Y., Wang, T., Chen, T., Sun, Z., Luo, D., Li, C., Wang, H., & Han, L. (2022). Evaluation of different machine learning models and novel deep learning-based algorithm for landslide susceptibility mapping. Geoscience Letters, 9, 26. https://doi.org/10.1186/s40562-022-00236-9

[4] Merghadi, A., Yunus, A. P., Dou, J., Whiteley, J., ThaiPham, B., Bui, D. T., Avtar, R., & Abderrahmane, B. (2020). Machine learning methods for landslide susceptibility studies: A comparative overview of algorithm performance. Earth-Science Reviews, 207, 103225. https://doi.org/10.1016/j.earscirev.2020.103225

[5] Kalantar, B., Ueda, N., Saeidi, V., Ahmadi, K., Halin, A. A., & Shabani, F. (2020). Landslide susceptibility mapping: Machine and ensemble learning based on remote sensing big data. Remote Sensing, 12(11), 1737. https://doi.org/10.3390/rs12111737

[6] Chen, Y., Wei, Y., Wang, Q., Chen, F., Lu, C., & Lei, S. (2020). Mapping post-earthquake landslide susceptibility: A U-Net like approach. Remote Sensing, 12(17), 2767. https://doi.org/10.3390/rs12172767

[7] Nachappa, T. G., Kienberger, S., Meena, S. R., Hölbling, D., & Blaschke, T. (2020). Comparison and validation of per-pixel and object-based approaches for landslide susceptibility mapping. Geomatics, Natural Hazards and Risk, 11(1), 572-600. https://doi.org/10.1080/19475705.2020.1736190

[8] Nhu, V.-H., Mohammadi, A., Shahabi, H., Ahmad, B. B., Al-Ansari, N., Shirzadi, A., Clague, J. J., Jaafari, A., Chen, W., & Nguyen, H. (2020). Landslide susceptibility mapping using machine learning algorithms and remote sensing data in a tropical environment. International Journal of Environmental Research and Public Health, 17(14), 4933. https://doi.org/10.3390/ijerph17144933

[9] Lee, D. H., Kim, Y. T., & Lee, S. R. (2020). Shallow landslide susceptibility models based on artificial neural networks considering the factor selection method and various non-linear activation functions. Remote Sensing, 12(7), 1194. https://doi.org/10.3390/rs12071194

[10] Karakas, G., Can, R., Kocaman, S., Nefeslioglu, H. A., & Gokceoglu, C. (2020). Landslide susceptibility mapping with random forest model for Ordu, Turkey. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLIII-B3-2020, 1229-1236. https://doi.org/10.5194/isprs-archives-XLIII-B3-2020-1229-2020

[11] Phong, T. V., Ly, H.-B., Trinh, P. T., Prakash, I., & Hoan, D. T. (2020). Landslide susceptibility mapping using Forest by Penalizing Attributes (FPA) algorithm based machine learning approach. Vietnam Journal of Earth Sciences, 42(3), 237-246. https://doi.org/10.15625/0866-7187/42/3/15047

[12] Xiao, T., Segoni, S., Chen, L., Yin, K., & Casagli, N. (2020). A step beyond landslide susceptibility maps: A simple method to investigate and explain the different outcomes obtained by different approaches. Landslides, 17, 627-640. https://doi.org/10.1007/s10346-019-01299-0

[13] Fang, Z., Wang, Y., Peng, L., & Hong, H. (2021). A comparative study of heterogeneous ensemble-learning techniques for landslide susceptibility mapping. International Journal of Geographical Information Science, 35(2), 321-347. https://doi.org/10.1080/13658816.2020.1808897

[14] Zhou, X., Wu, W., Qin, Y., & Fu, X. (2021). Geoinformation-based landslide susceptibility mapping in subtropical area. Scientific Reports, 11, 24325. https://doi.org/10.1038/s41598-021-03743-5

[15] Huang, F., Cao, Z., Guo, J., Jiang, S.-H., Li, S., & Guo, Z. (2020). Comparisons of heuristic, general statistical and machine learning models for landslide susceptibility prediction and mapping. Catena, 191, 104580. https://doi.org/10.1016/j.catena.2020.104580

[16] Huang, F., Yin, K., Huang, J., Gui, L., & Wang, P. (2020). Landslide susceptibility prediction based on a semi-supervised multiple-layer perceptron model. Landslides, 17(12), 2919-2930. https://doi.org/10.1007/s10346-020-01473-9

[17] Panahi, M., Gayen, A., Pourghasemi, H. R., Rezaie, F., & Lee, S. (2020). Spatial prediction of landslide susceptibility using hybrid support vector regression and the adaptive neuro-fuzzy inference system with various metaheuristic algorithms. Science of the Total Environment, 741, 139937. https://doi.org/10.1016/j.scitotenv.2020.139937

[18] Pham, B. T., Prakash, I., Dou, J., Singh, S. K., Trinh, P. T., Tran, H. T., Le, T. M., Van Phong, T., Khoi, D. K., & Bui, D. T. (2020). GIS-based ensemble soft computing models for landslide susceptibility mapping. Advances in Space Research, 66(6), 1303-1320. https://doi.org/10.1016/j.asr.2020.05.016

[19] Zhu, L., Huang, Y., Fan, L., Huang, J., Huang, F., Chen, J., & Wang, Y. (2020). Landslide susceptibility prediction modeling based on remote sensing and a novel deep learning algorithm of a cascade-parallel recurrent neural network. Sensors, 20(6), 1576. https://doi.org/10.3390/s20061576

[20] Zhang, T.-Y., Mao, Z.-A., & Wang, T. (2020). GIS-based evaluation of landslide susceptibility using a novel hybrid computational intelligence model on different mapping units. Journal of Mountain Science, 17(12), 2929-2941. https://doi.org/10.1007/s11629-020-6393-8

[21] Li, W., Wang, D., Li, H., & Chen, W. (2020). Uncertainties analysis of collapse susceptibility prediction based on remote sensing and GIS: Influences of different data-based models and connections between collapses and environmental factors. Remote Sensing, 12(24), 4134. https://doi.org/10.3390/rs12244134

[22] Woodard, J. B., Mirus, B. B., Crawford, M., Or, D., Leshchinsky, B., Allstadt, K. E., & Wood, N. J. (2022). Mapping landslide susceptibility over large regions with limited data. Journal of Geophysical Research: Earth Surface, 127(12), e2022JF006810. https://doi.org/10.1029/2022JF006810

[23] Steger, S., Brenning, A., Bell, R., & Glade, T. (2021). The influence of systematically incomplete shallow landslide inventories on statistical susceptibility models and suggestions for improvements. Landslides, 18, 3269-3282. https://doi.org/10.1007/s10346-021-01693-2

[24] Wang, Y., Fang, Z., & Hong, H. (2020). Comparison of convolutional neural networks for landslide susceptibility mapping in Yanshan County, China. Science of the Total Environment, 710, 136315. https://doi.org/10.1016/j.scitotenv.2019.136315

[25] Azarafza, M., Azarafza, M., Akgün, H., Atkinson, P. M., & Derakhshani, R. (2021). Deep learning-based landslide susceptibility mapping. Scientific Reports, 11, 24112. https://doi.org/10.1038/s41598-021-03585-1

[26] Dornik, A., Drăguţ, L., Oguchi, T., Hayakawa, Y., & Micu, M. (2022). Influence of sampling design on landslide susceptibility modeling in lithologically heterogeneous areas. Scientific Reports, 12, 2106. https://doi.org/10.1038/s41598-022-06257-w