EEET ECOLOGICAL ENGINEERING & ENVIRONMENTAL TECHNOLOGY

Ecological Engineering & Environmental Technology, 2026, 27(8), 276–293
https://doi.org/10.12912/27197050/225243
ISSN 2719–7050, License CC-BY 4.0

Received: 2026.06.20
Accepted: 2026.07.10
Published: 2026.08.01

Random forest based spatial modeling of landslide susceptibility and land use exposure for sustainable land management in rural area

Muhammad Arya Farhan Rachman[1], Sumbangan Baja[2*], Samsu Arif[3]

[1] Environmental Management Study Program, The Graduate School, Hasanuddin University, Makassar 90245, Indonesia
[2] Department of Soil Science, Faculty of Agriculture, Hasanuddin University, Makassar 90245, Indonesia
[3] Department of Physics, Faculty of Mathematics and Natural Science, Hasanuddin University, Makassar 90245, Indonesia
* Corresponding author's e-mail: s.baja@unhas.ac.id

ABSTRACT

Landslides pose a severe environmental and socioeconomic threat in mountainous regions, yet existing spatial risk assessments often isolate hazard prediction from actual land-use exposure. This study develops an integrated geospatial framework to evaluate landslide susceptibility and quantify anthropogenic exposure in Tinggimoncong District, Indonesia, to support sustainable land management. A Random Forest algorithm was utilized within Google Earth Engine to model landslide susceptibility using an inventory of 198 landslide and non-landslide points and 11 parameter predictors. Concurrently, a land-use classification was performed using Sentinel-2A imagery. The susceptibility model achieved robust predictive performance with an area under the curve (AUC) of 0.843 and an overall accuracy of 80.3%, identifying slope gradient and distance to roads as the most dominant landslide-driving factors. The validated land-use map with overall accuracy of 85.6% was spatially intersected with high-susceptibility zones (>0.7), revealing that 16.97% of the road infrastructure and 4.90% of settlements are highly exposed to landslide hazards, contrasting with a minimal exposure 0.22% for agricultural lands. These findings provide a robust scientific foundation for sustainable land management, recommending targeted interventions such as a moratorium on new construction in red zones, structural geotechnical engineering on vulnerable road-cut slopes, and deep-rooted vegetative conservation on exposed agricultural lands.

Keywords: landslide susceptibility model, random forest, spatial exposure, Sentinel-2A, sustainable land management.

INTRODUCTION

Landslides are a serious threat with destructive impacts on the physical environment, infrastructure, and social life of communities. Landslide susceptibility can increase drastically due to the complex interactions between climate change, land-use changes, and population growth pressures in previously stable areas (Duan et al., 2025). Tinggimoncong District in Gowa Regency is an area with an extreme level of landslide susceptibility. Historically, this region experienced a massive landslide at Mount Bawakaraeng in 2004, triggering a debris flow with a volume reaching 232 million cubic meters (Tsuchiya et al., 2004). The disaster, triggered by the collapse of the caldera wall, resulted in severe environmental degradation, the destruction of various facilities, and claimed 32 lives. This threat is persistent; in 2022, the National Disaster Management Agency (BNPB) recorded further road infrastructure damage and casualties due to landslides triggered by high-intensity rainfall on the connecting route between Tinggimoncong and Parangloe. The high intensity of these disasters is also confirmed by records of recurrent events from the South Sulawesi Regional Disaster Management Agency (BPBD) throughout early 2025.

To effectively mitigate landslide disasters, accurate spatial information regarding the distribution of susceptible areas is required. Recent publications demonstrate a growing shift from conventional heuristic approaches toward data-driven machine learning frameworks for environmental hazard prediction and spatial risk assessment Dolchinkov (2024). Researchers successfully integrate machine learning methods with remote sensing data and spatial environmental datasets to improve the analytical capabilities of geospatial assessment systems (Ejaz et al., 2026). Literature emphasizes that modern landslide risk modeling increasingly integrates biophysical parameters with land-use and governance-driven anthropogenic factors to better represent real-world susceptibility dynamics (Millimono, 2026). Recent studies confirm that Random Forest-based models are highly effective for landslide susceptibility mapping in complex mountainous environments using multi-factor geospatial datasets including slope, rainfall, land use, and anthropogenic disturbances (Firoozi et al., 2024). Conventional landslide susceptibility mapping frequently employs weighted overlay techniques; however, this approach has limitations as it relies heavily on subjective assumptions or expert opinions (Chowdhury, 2023). To overcome this subjective bias, the utilization of advanced computational instruments such as Machine Learning is highly recommended. Among the various available algorithms, Random Forest has been empirically proven capable of unraveling the complexity of slope biophysical parameters and generating susceptibility mapping with superior accuracy compared to other machine learning methods (Akinci et al., 2020; Merghadi et al., 2020).

Despite the growing application of machine learning methods in landslide susceptibility mapping, most existing studies primarily focus on hazard prediction alone, without explicitly linking susceptibility outputs to real-world land-use exposure and spatial planning implications. In addition, the majority of previous works rely on susceptibility maps as standalone products, without quantifying how different categories of anthropogenic land use are spatially affected within high-risk zones, particularly in rapidly developing mountainous regions such as Tinggimoncong District. This creates a critical gap between predictive hazard modeling and actionable spatial decision-making.

Furthermore, there is limited methodological integration between machine learning-based susceptibility modeling and remote sensing-derived land-use classification within a unified spatial framework that enables direct evaluation of exposure patterns. As a result, current approaches often fail to translate model outputs into practical land management priorities, especially in areas characterized by complex topography and high infrastructure vulnerability.

To address this gap, this study hypothesizes that coupling Random Forest-based landslide susceptibility modeling with high-resolution satellite-derived land-use classification can significantly improve the interpretability and policy relevance of spatial risk assessments by explicitly identifying the distribution of exposed land-use types within high-susceptibility zones.

Accordingly, the main objective of this research is to develop an integrated geospatial framework that (i) models landslide susceptibility using a Random Forest algorithm, (ii) classifies land-use patterns using Sentinel-2 imagery, and (iii) quantifies spatial exposure through overlay analysis. The expected scientific contribution is not only the generation of a susceptibility map with high predictive performance, but also the establishment of a reproducible workflow that links hazard probability to land-use exposure, thereby improving the applicability of machine learning outputs for spatial planning and disaster risk reduction in mountainous environments.

MATERIALS AND METHODS

Study area

This research was conducted in Tinggimoncong District, Gowa Regency, South Sulawesi Province, Indonesia. Geographically, the study area is situated between ±5°11’ and 5°18’ South Latitude, and ±119°46’ to 119°56’ East Longitude, covering a total area of 167.4 km^2. The region is characterized by a complex and steep mountainous topography, with elevations ranging from 260 to 2880 meters above sea level and slope angles reaching up to 64°. Furthermore, the area experiences high annual rainfall intensity, ranging from 2700 to 3100 mm/year, which significantly contributes to its extreme susceptibility to landslide events (Figure 1).

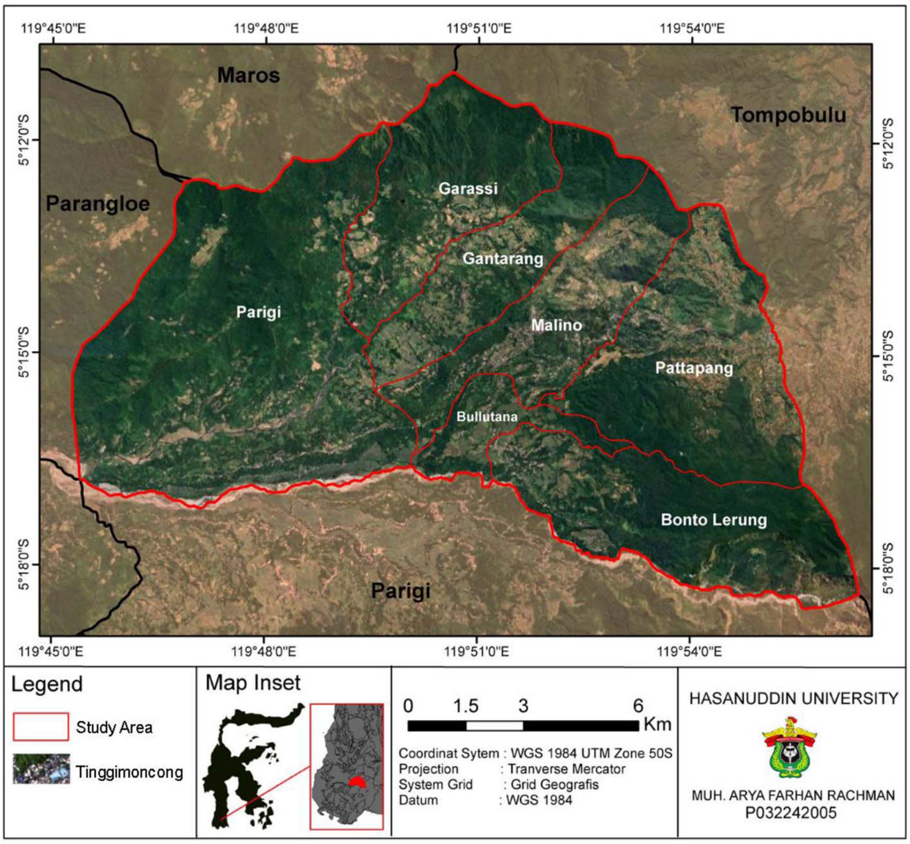

Figure 1. Study area map

Data preparation

Landslide inventory source was from direct field surveys combined with remote sensing visual interpretation using Google Earth Pro, covering the 2014–2025 period, comprising a total of 99 landslide points.

Non-landslide sampling method was random sampling using a buffer distance approach, maintaining a strictly balanced 1:1 ratio (99 non-landslide points).

Spatial resolution of all parameter raster layers of 30 m – all predictor parameter/variables were resampled to 30 m for the Random Forest modeling to match the native DEM resolution, while the land-use map was maintained at 10 m for the final high-resolution exposure assessment.

Projection system used WGS 1984 UTM Zone 50S (EPSG: 32750).

Landslide susceptibility mapping

Landslide susceptibility was modeled using a Random Forest (RF) classifier implemented within the Google Earth Engine (GEE) JavaScript API. The input dataset consisted of a total of 198 points, comprising 99 historical landslide locations and 99 non-landslide samples, maintaining a rigorously balanced 1:1 ratio. The landslide inventory was acquired through a combination of direct field surveys and historical visual interpretation utilizing Google Earth Pro covering the period from 2014 to 2025. The inclusion criteria for the inventory were strictly limited to mass movements exhibiting clearly identifiable main scarps or significant accumulations of displaced material. Conversely, minor surface erosions were explicitly excluded to ensure the dataset accurately represents true slope failure events. The non-landslide samples were generated using a random buffer distance approach to ensure distinct environmental and spatial separation from actual landslide occurrences.

The predictor variables included slope, slope aspect, slope length, elevation, lithology, soil type, rainfall intensity, distance to stream, distance to road, flow accumulation, and land use. Because tree-based algorithms like RF are inherently invariant to monotonic transformations and feature scaling, all environmental predictors were kept unstandardized. The RF model was trained using the following optimized hyperparameters: 500 decision trees (*numberOfTrees*), an unrestricted maximum tree depth, a minimum of 1 sample per leaf (*minLeafPopulation*), and 3 features randomly selected per split (*variablesPerSplit*). Each decision tree was constructed using bootstrap samples drawn with replacement (bag fraction: 0.632), with node splitting executed based on the Gini impurity criterion. Model aggregation was performed using a probability output mode across the ensemble of 500 trees. To eliminate spatial data leakage and rigorously evaluate the model's generalization capacity against unseen data, the random train-test split was replaced with a 10-fold spatial block cross-validation. The study area was divided into spatially independent 2.5 × 2.5 km grids.

The relative contribution of each predictor variable to the landslide susceptibility model was evaluated using both feature importance and permutation importance. These two approaches were presented to provide complementary perspectives on variable relevance. Feature importance quantifies the contribution of each predictor during the model construction process by measuring how frequently and effectively a variable is used to split decision trees. In contrast, permutation importance evaluates the actual dependence of the trained model on each predictor by measuring the decrease in prediction accuracy after the values of that variable are randomly permuted. Consequently, the combined use of these methods enables a more comprehensive interpretation of predictor importance, allowing the distinction between variables that contribute substantially to model construction and those that are genuinely essential for maintaining predictive performance.

Model performance was evaluated using a global confusion matrix to calculate overall accuracy, producer's accuracy (recall), consumer's accuracy (precision), Kappa coefficient, and the F1-score directly within the GEE environment. Furthermore, to address continuous-variable bias and rigorously evaluate discriminative capability, the receiver operating characteristic - area under the curve (ROC-AUC) and permutation feature importance were externally computed using Python 3.x and the *scikit-learn* library via Google Colab.

To identify regions with high ambiguity, an uncertainty map was generated using the Shannon Entropy index. Since the Random Forest model outputs a continuous probability of landslide occurrence, pixel-level uncertainty was evaluated directly based on this probability distribution. In this framework, an entropy value approaching 1 indicates maximum uncertainty where the predicted probability is near 0.5 signifying high model ambiguity due to conflicting predictions among the individual decision trees. Conversely, an entropy value approaching 0 indicates a very high level of prediction confidence, reflecting uniform decisions across the decision tree ensemble.

The random forest algorithm

The Random Forest algorithm operates by constructing an ensemble of multiple decision trees. Each decision tree is trained independently using the Bootstrap technique, which involves random sampling of parameters with replacement.The input parameters used in this study were selected based on factors that have been empirically demonstrated to influence landslide occurrence. Landslide susceptibility is controlled by a combination of geological, topographical, and anthropogenic factors, with soil type, soil erodibility, and lithology playing significant roles in machine learning-based susceptibility modeling. Topographic characteristics, particularly slope gradient, are widely recognized as key factors affecting slope stability. In addition, precipitation patterns strongly influence soil moisture conditions and contribute to the initiation of landslide events (Wu et al., 2025). Land use represents a highly dynamic factor that reflects human-induced landscape changes and has been identified as an important variable in landslide susceptibility assessment (Rabby et al., 2022). To further improve model performance, additional spatial parameters, including elevation, distance to rivers, and distance to roads, were incorporated. Elevation reflects geomorphological variability across the landscape, while proximity to rivers and roads is associated with slope modification, erosion processes, and anthropogenic disturbances. As reported by Zandi et al. (2025), landslide occurrence is frequently concentrated in areas affected by river incision and road construction, where natural slope equilibrium has been substantially altered.

In the classification process, node splitting in each tree does not evaluate all landslide triggering parameters simultaneously. Instead, it utilizes a random feature selection formulated as $m = \sqrt{M}$ where m represents the number of randomly selected features at each split and M denotes the total number of parameters used. This formula ensures variation among the trees and prevents any single parameter from dominating others, thereby enhancing the diversity of the decision trees (Chen et al., 2025). The determination of the best landslide-determining parameters is mathematically calculated using Gini impurity to measure how effectively a parameter classifies landslide susceptibility areas based on landslide and non-landslide points.

$$Gini\ impurity = 1 - \sum_{i=1}^{c} p_i^2 \quad (1)$$

where: C – number of classes, p_i – proportion of samples in the node belonging to class i.

Subsequently, the effectiveness of the parameters in classifying landslide susceptibility areas is calculated using information gain, which represents the magnitude of the reduction in Gini impurity. A lower Gini impurity value yields a higher information gain, indicating that the parameter is more optimal for the modeling process (Yao et al., 2020).

$$Information\ gain = I(parent) - \\ - \sum_{j=1}^{k} \frac{N_j}{N} I(child_j) \quad (2)$$

where: $I(parent)$ – Gini impurity value of the parent node, $I(child)$ – Gini impurity value of the j-th child node (new label split), N_j – number of labels/samples in the j-th child node, N – total number of labels/samples in the parent node, k – number of child nodes after the split.

The probability output of the Random Forest algorithm was utilized to generate a continuous landslide susceptibility map. The final susceptibility index for each map pixel, which ranges from 0 to 1, is calculated based on the proportion of decision trees in the ensemble that predict the landslide-occurrence class. During the training phase, the model internally evaluates its prediction error using out-of-bag (OOB) data, which consists of the samples not selected during the Bootstrap process.

$$OOB_{error} = \frac{1}{N} \sum_{i=1}^{N} I\left(y_i \neq \hat{y}_i^{OOB}\right) \quad (3)$$

where: I – indicator function (value is 1 if the prediction is incorrect, 0 if the prediction is correct), $\hat{y}_i$ – OOB prediction for the i-th sample, y_i – actual class for the i-th sample, N – number of data points in the training set.

Finally, to ensure the reliability and geographical robustness of the predictive model, external validation was rigorously conducted using a Spatial 10-fold cross-validation approach. Unlike random splitting, this spatial partitioning method mitigates spatial autocorrelation by dividing the study area into mutually exclusive geographical grids. The model's predictive performance across these spatial folds was comprehensively quantified using a confusion matrix to derive overall accuracy, kappa coefficient, and F1-score, alongside the receiver operating characteristic (ROC) curve to evaluate the area under the curve (AUC) (Çorbacıoğlu and Aksel, 2023; Wang et al., 2024).

Land-use classification

Land-use classification was performed using Sentinel-2A multispectral imagery with 10 m spatial resolution, which offers superior spatial and spectral capabilities suitable for tropical land cover mapping compared to medium-resolution predecessors (Phiri et al., 2020). The image acquisition date used in this study was September 5, 2025. The following spectral bands were used for classification: B2, B3, B4, B8, B11, and B12. Prior to classification, atmospheric and radiometric preprocessing applied using the Copernicus Level-2A (surface reflectance) product, and the selected bands were stacked and resampled to a uniform spatial resolution of 10 m.

Supervised classification was conducted using a RF algorithm implemented natively within the Google Earth Engine (GEE) environment, an approach selected to capture complex spatial dependencies more effectively than traditional static methods (Hussein et al., 2025). The classifier was trained using regions of interest (ROI) manually digitized to serve as representative training examples (Yulianto et al., 2021). In this study, the ROIs were specifically delineated for five land-use classes: settlement, water bodies, wetland agriculture, dryland agriculture, and forest.

To ensure class balance and mitigate prediction bias, the number of training samples (polygons) was proportionally distributed across the landscape as follows: settlement (185), water bodies (115), wetland agriculture (145), dryland agriculture (120), and forest (110). The sample regions function converted the ROI polygons into pixel-based samples at 10 m spatial resolution. The dataset was split into training and validation samples using a stratified random sampling approach with a 70:30 ratio (70% training and 30% testing) and a fixed random seed (seed = 42) to ensure reproducibility.The Random Forest model was configured with 100 trees, while the remaining parameters were kept at their default settings. Classification was performed on pixel-level spectral input data.

To reduce misclassification between spectrally similar classes, particularly water bodies and built-up areas, post-classification refinement was applied using raster patching, implemented as a manual post-classification GIS-based reclassification procedure. Areas exhibiting clear classification inconsistencies were identified through visual comparison with very high-resolution Google Earth imagery. Only the identified anomalous patches were corrected using the reclassify tool, while the original Random Forest classification was preserved for the remaining study area. The corrected patches were subsequently merged with the original classified raster using the mosaic to new raster function with the last mosaic operator, ensuring that the corrected raster replaced the corresponding erroneous classifications.

The resulting land-use map was validated using the allocated 30% testing subset (N = 5665 pixels). Interpretation of the land-use classes was supported by very high-resolution Google Earth imagery and Sentinel-2 true-color and false-color composites. Validation was performed using a confusion matrix, from which overall accuracy (OA) and kappa coefficient were computed. Overall Accuracy was calculated as:

$$OA = \left(\frac{\sum x_{ii}}{N}\right) \times 100\% \quad (4)$$

where: x_{ii} represents correctly classified pixels and N represents total validation samples.

The Kappa coefficient was calculated using the standard Cohen's Kappa equation implemented within the Google Earth Engine environment to account for agreement beyond chance. All spatial processing was performed in ArcGIS 10.8 using coordinate reference system WGS 1984 UTM Zone 50S (EPSG: 32750). This evaluation demonstrates the precision and reliability of the map, validating its suitability as an input for subsequent spatial overlay analysis with the landslide susceptibility map (Bayas et al., 2022).

Land-use exposure

Land-use exposure analysis was conducted through a pixel-based spatial overlay between the landslide susceptibility raster and the classified land-use map. Both datasets were resampled to a common spatial resolution of 10 meters and aligned to the same coordinate reference system (EPSG: 32750, representing WGS 1984 UTM Zone 50S). Landslide susceptibility values were first categorized into discrete classes using fixed thresholds - Equal Interval, the high susceptibility areas were defined as pixels with landslide susceptibility values more than 0.7 (De Praga Baião et al., 2025; Ganesh et al., 2025; Xing et al., 2026)

The exposure analysis was implemented using ArcGIS 10.8 by performing a raster overlay operation, where land-use classes were intersected with the high-susceptibility mask. The output was a categorical exposure map indicating spatial coincidence between each land-use class and landslide susceptibility zones. Exposure statistics were computed as the total area of each land-use class within each susceptibility category. Area calculations were performed by multiplying the pixel count by the pixel area using the GIS zonal statistics tool. For linear features, road infrastructure was represented as vector data; spatial intersection between vector roads and raster susceptibility zones was performed using a buffer distance and raster overlay approach with a buffer distance of 5 meters.

All spatial outputs were generated in ArcGIS 10.8 using the projection system EPSG: 32750. The final exposure map was used to identify spatial hotspots where anthropogenic assets (such as built-up areas, agricultural land, and road networks) coincide with high landslide susceptibility. Furthermore, the proposed framework supports the transition from generalized disaster management toward more data-driven and site-specific strategies aimed at enhancing resilience in rural mountainous regions

RESULTS AND DISCUSSION

Landslide susceptibility map

Landslide susceptibility mapping in Tinggimoncong District was successfully conducted using the Random Forest machine learning algorithm. The landslide susceptibility map was generated by training the model with a landslide inventory dataset consisting of landslide and non-landslide points. The landslide points were collected through direct field observations and satellite image interpretation, as illustrated in Figures 2

Figure 2. Direct field inventory of landslide points

and 3, while their spatial distribution is presented in Figure 4. In addition to the landslide inventory, multiple landslide conditioning factors were incorporated into the modeling process. These conditioning factors are presented in Figure 5.

Figure 4 illustrates the spatial distribution of landslide and non-landslide samples together with the spatial blocks used in this study. The study area was first divided into 2.5 × 2.5 km spatial grid cells. Each grid cell was randomly assigned to one of ten folds. During each iteration of the 10-fold spatial block cross-validation, all samples located within the selected fold were used as the validation dataset, while samples in the remaining folds were used for model training. This strategy ensures spatial independence between the training and validation datasets and reduces the influence of spatial autocorrelation.

Each parameter used in this study contributed differently to landslide occurrence. The parameter selection strategy is consistent with the machine learning-based selective measuring framework proposed by Baja et al. (2025), which suggests that environmental modeling can be conducted more effectively and efficiently by focusing computational resources and field observations on the parameters with the highest predictive value. Therefore, it is necessary to evaluate the importance of each parameter to determine its contribution to model construction and its influence on predictive performance.

The evaluation was conducted using two metrics, namely feature importance and permutation importance. Although both metrics aim to identify the most influential parameters, they are based on different calculation principles. Feature

Figure 3. Landslide point inventory using satellite imagery

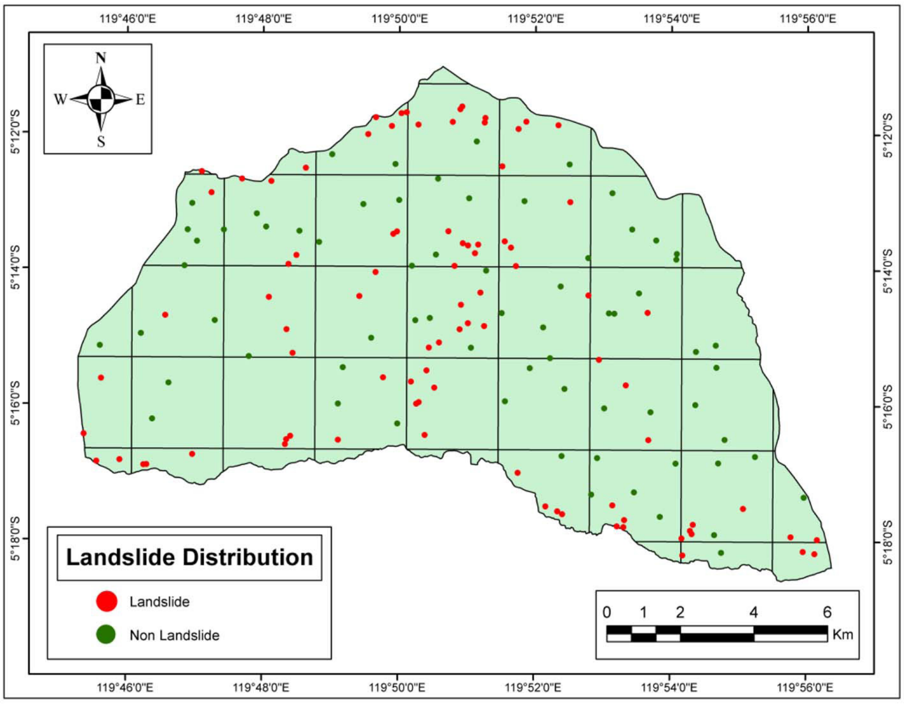

Figure 4. Landslide distribution map

Importance reflects the contribution of each parameter during the model construction process, whereas Permutation Importance measures the model's dependence on each parameter by evaluating the decrease in predictive performance when the values of that parameter are randomly permuted. Therefore, presenting both metrics together provides a more comprehensive understanding of the role of each parameter, both in model construction and in maintaining the model's predictive performance. The resulting values were then used to rank the parameters according to their relative importance, as shown in Figure 6.

Based on Table 1, most parameters exhibited relatively consistent importance rankings between feature importance and permutation importance. Slope and distance to road consistently ranked first and second, indicating that these two parameters were the most dominant factors both in the model construction process and in maintaining predictive performance. Elevation, slope aspect, distance to stream, slope length, flow accumulation, land use, and soil type also showed only minor or no changes in their rankings, suggesting that these parameters made stable contributions to the model. Rain intensity exhibited a high feature importance but a lower permutation importance, indicating that the information it provides is partially redundant with that of other parameters. In contrast, Lithology showed relatively low importance according to feature importance but considerably higher importance according to permutation importance. This finding suggests that lithological information was not frequently used to create decision tree splits, despite being important for maintaining the model's predictive accuracy. This is likely because Lithology is a categorical variable; according to Vodde et al. (2025), continuous variables are generally utilized more frequently by the Random Forest algorithm than categorical variables.

Table 1. Comparison of feature importance and permutation importance

Rank	Feature Importance	Permutation Importance
1	Slope	Slope
2	Distance to road	Distance to road
3	Elevation	Lithology
4	Slope aspect	Elevation
5	Rain intensity	Distance to stream
6	Distance to stream	Slope aspect
7	Slope length	Slope length
8	Flow accumulation	Flow accumulation
9	Lithology	Rain Intensity
10	Land use	Land use
11	Soil type	Soil type

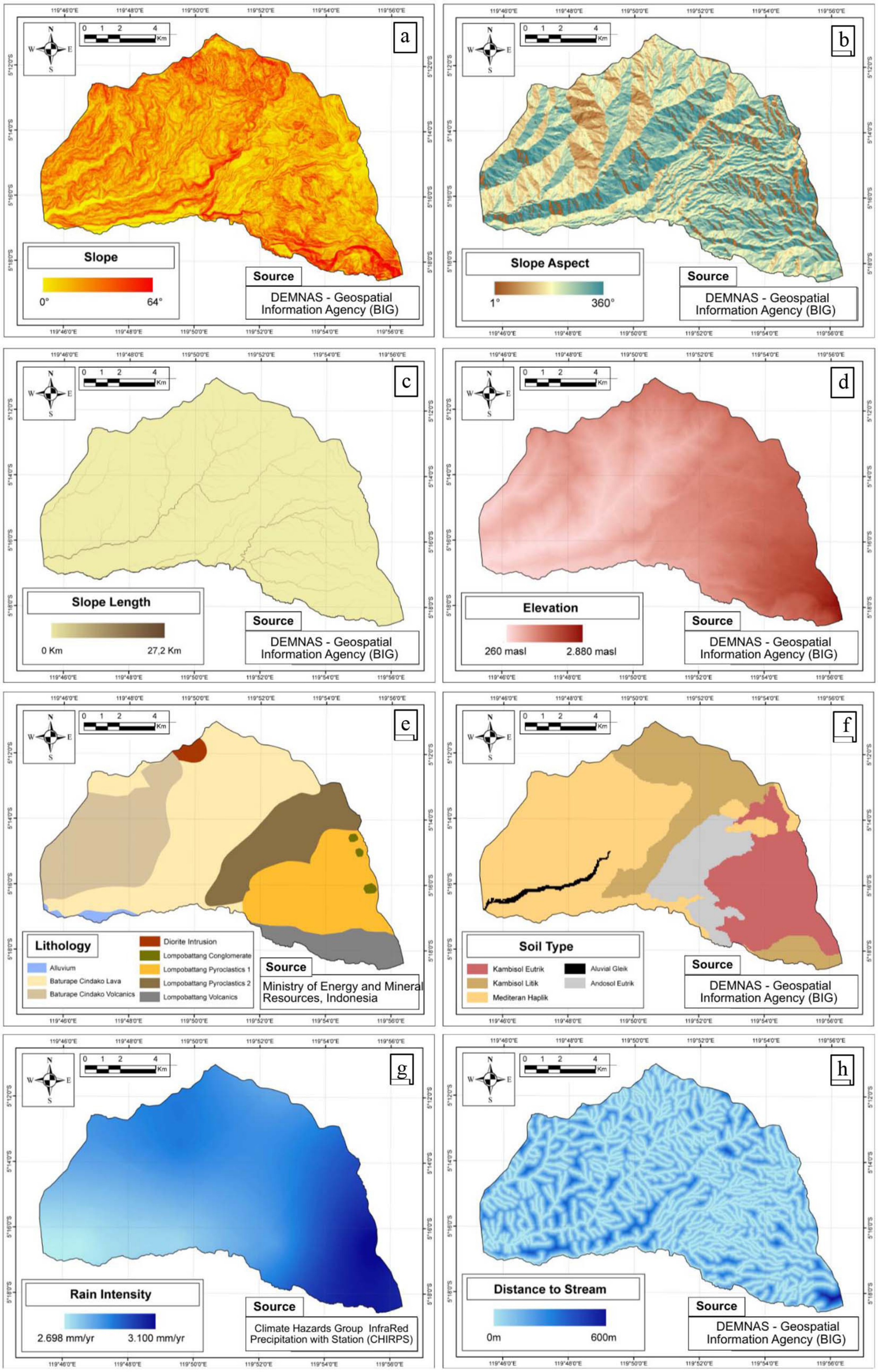

Figure 5. Landslide parameters: (a) slope, (b) slope aspect, (c) slope length, (d) elevation, (e) lithology, (f) soil type, (g) rain intensity, (h) distance to rivers, (i) distance to roads, (j) flow accumulation, (k) land use

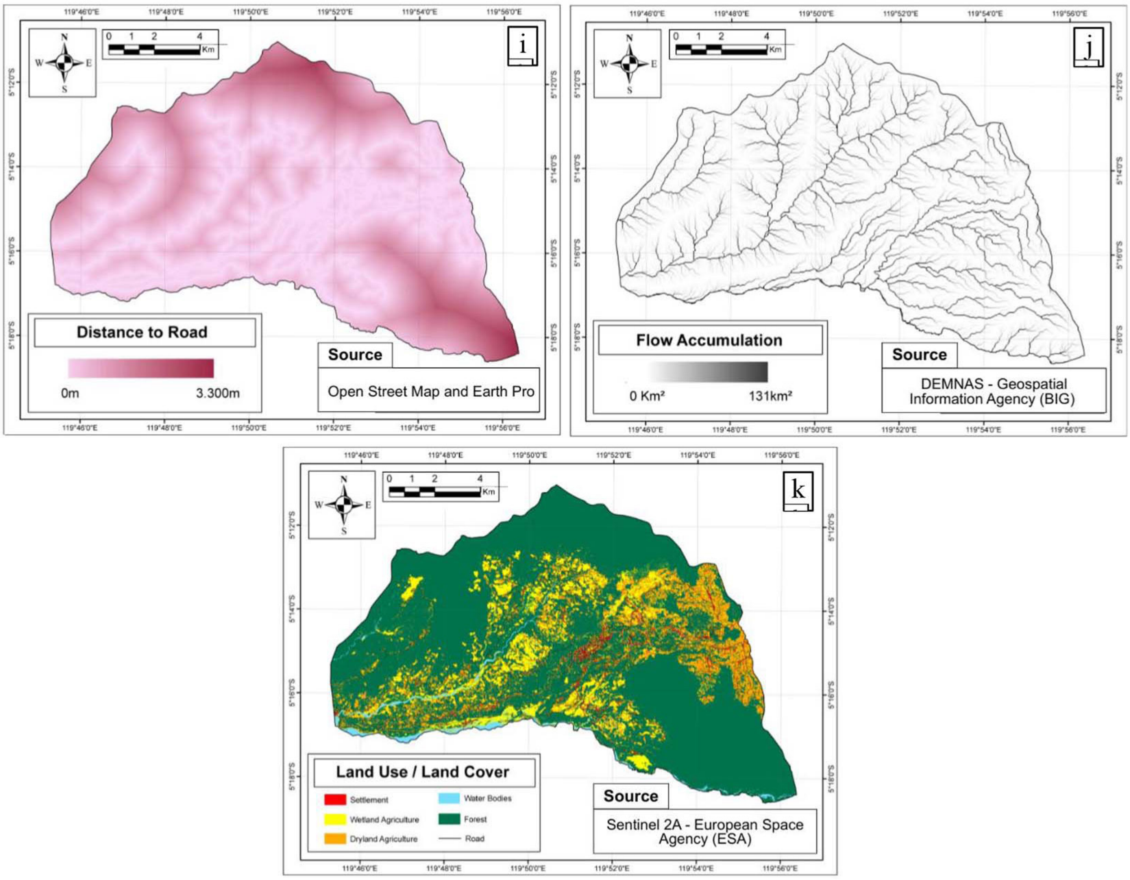

Figure 5. Cont. Landslide parameters: (a) slope, (b) slope aspect, (c) slope length, (d) elevation, (e) lithology, (f) soil type, (g) rain intensity, (h) distance to rivers, (i) distance to roads, (j) flow accumulation, (k) land use

Figure 7 presents the landslide susceptibility map, which displays landslide occurrence probability values ranging from 0 to 1. This indicates that the closer the value is to 1, the greater the likelihood of a landslide event occurring.

The receiver operating characteristic (ROC) curve shown in Figure 8 yielded an area under the curve (AUC) value of 0.843, indicating good classification performance. This value demonstrates that the model has a high discriminative ability to assign higher probabilities to locations where landslides have occurred than to locations where no landslides have occurred.

Based on the confusion matrix in Table 2, the model successfully classified 81 non-landslide samples as true negatives (TN) and 78 landslide samples as true positives (TP), while the misclassifications consisted of 19 false positives (FP) and 20 false negatives (FN). From these figures, relatively balanced precision, recall, and F1-score values were generated for both classes. The precision value of 0.802 for the non-landslide class and 0.804 for the landslide class indicates that approximately 80% of the generated predictions are correct classifications. Meanwhile, a recall value of 0.810 for the non-landslide class and 0.796 for the landslide class indicates that the model is capable of identifying around 80% of the actual occurrences in each class. This balance is reinforced by the F1-scores of 0.806 and 0.800, which demonstrate that the model is able to maintain a balance between prediction accuracy and the ability to detect occurrences across both classes (Zhang et al., 2024). The very small differences in values among the metrics, along with the nearly equal number of False positives and false negatives, indicate that the model does not exhibit any tendency or bias toward either class, resulting in consistent predictions.

Furthermore, the overall accuracy is 80.3%, which indicates that eight out of every ten samples were predicted correctly. Additionally, the model achieved a Kappa coefficient (or Cohen's Kappa) of 0.606, indicating that the level of agreement between the prediction results and the actual data falls into the good or "substantial agreement" category. Although various evaluation metrics show

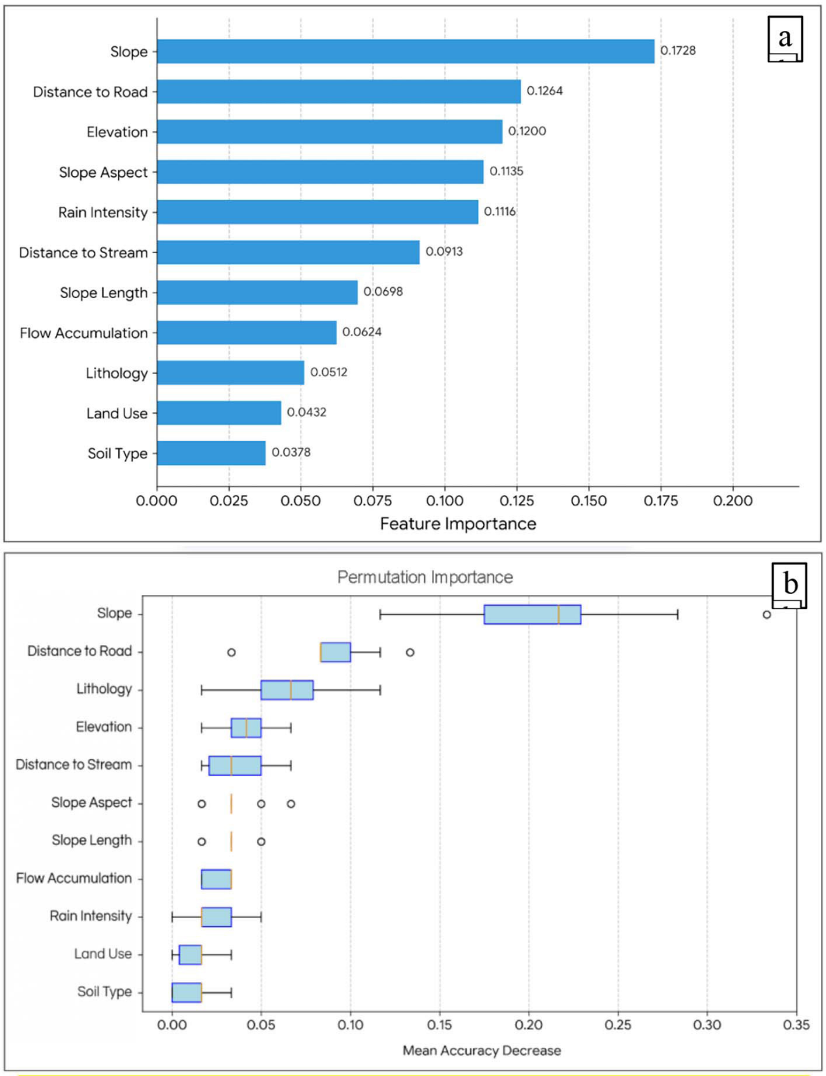

Figure 6. (a) Feature importance, (b) permutation importance

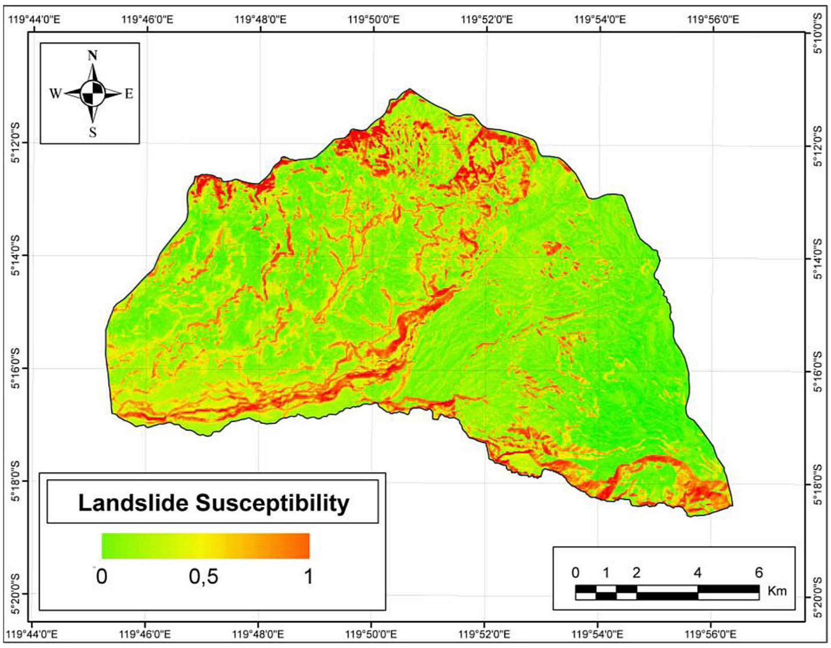

Figure 7. Landslide susceptibility map

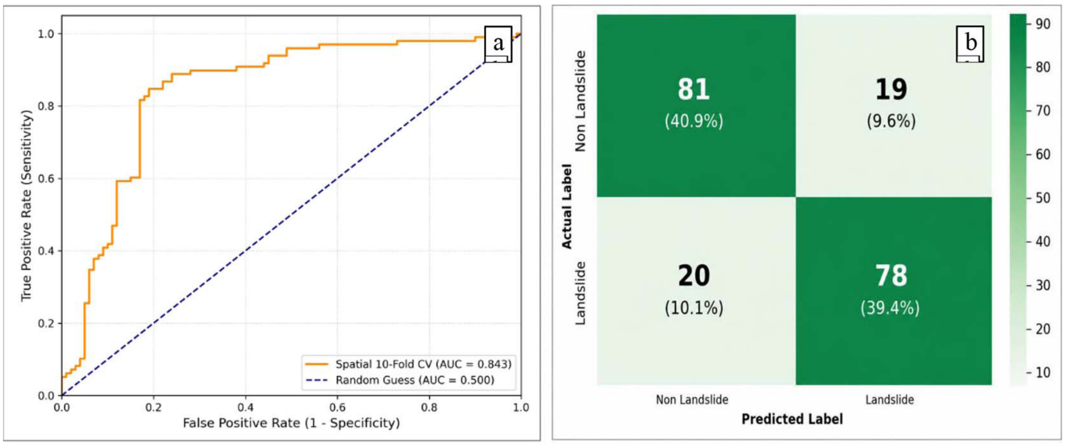

Figure 8. (a) ROC curve and (b) confussion matrix

Table 2. Confusion matrix of the landslide susceptibility modeling

Actual condition	Predicted condition	
	Non-landslide	Landslide
Non-landslide	81 TN	19 FP
Landslide	20 FN	78 TP
Metric	Value	
	Non-landslide	Landslide
Precision	0.80198	0.80412
Recall	0.81000	0.79592
F1-score	0.80597	0.80000
Overall accuracy	0.803	
Kappa coefficient	0.606	

that the Random Forest model has good and consistent classification performance, these evaluations only depict the overall performance of the model and do not yet reveal the spatial prediction confidence level at each location. The spatial prediction confidence level can be seen in Figure 9.

The uncertainty map generated using Shannon Entropy to visualize the level of model prediction confidence at each pixel. Conceptually, a low entropy value reflects the dominance of the probability of a single class (landslide or non-landslide) within a pixel, whereas a high value indicates model ambiguity due to balanced class probabilities. Spatially, the map demonstrates that

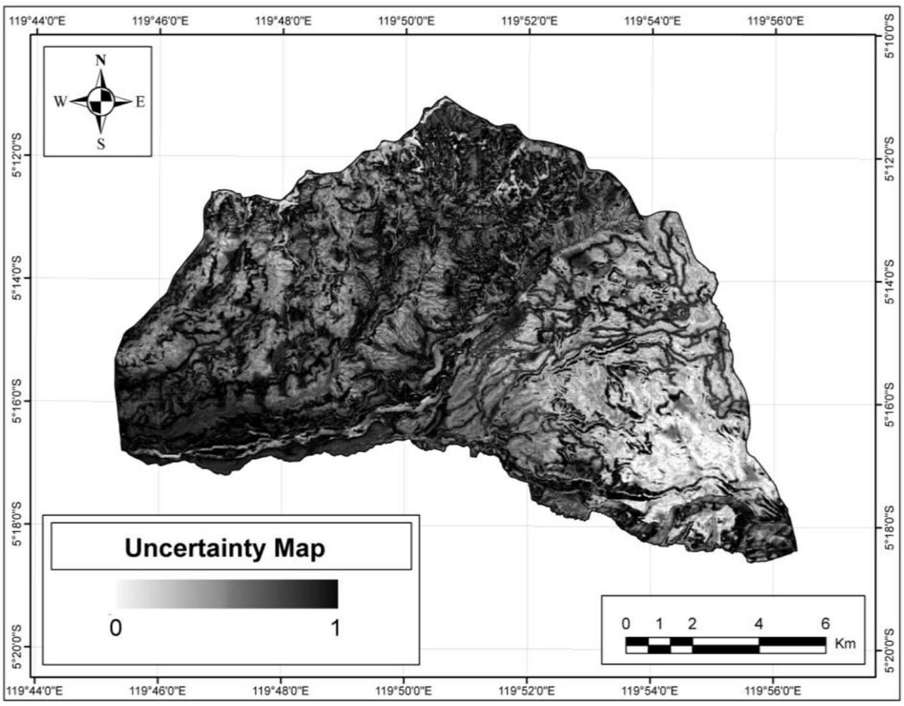

Figure 9. Uncertainty map

a large portion of the study area is dominated by low entropy values, indicating uniform prediction agreement within the decision tree ensemble. Conversely, areas with high uncertainty tend to be concentrated in linear patterns strongly associated with road corridors and zones of sharp topographic changes. This pattern confirms that slope modifications in these locations create complex environmental characteristics, making it difficult for the model to decisively separate the two classes.

Nevertheless, the high predictive uncertainty in these specific locations does not diminish the model's reliability in determining mitigation priorities. This is because maximum uncertainty (an entropy value approaching 1) conceptually only occurs when the prediction probability is within a moderate range, around 0.5. In contrast, the land-use exposure analysis in this study specifically extracted susceptibility zones using a high threshold of >0.7. Through this thresholding strategy, the areas identified for mitigation priority automatically bypass the zones of model ambiguity and possess a very high level of prediction confidence. This approach ensures that the identified hazard hotspots remain valid and robust to serve as a scientific foundation for spatial planning and decision-making.

Land-use classes

The land use classification in this study was performed using the Random Forest algorithm, which was selected for its ability to model nonlinear relationships among spectral variables and achieve relatively high classification accuracy. The initial Random Forest classification was able to represent the overall land use conditions; however, several misclassifications were still identified in specific locations. These errors primarily occurred between the water body and settlement classes due to the similarity of their spectral reflectance values, making them difficult for the model to distinguish. Therefore, a raster patching procedure was applied as a post-classification refinement step to correct the misclassified areas while preserving the original classification results in the remaining regions (Figure 10).

By preserving the original Random Forest classification across most of the study area and correcting only the areas that exhibited clear classification errors, the spatial quality of the land use map became more consistent with actual ground conditions. This refinement resulted in an improvement in classification accuracy. The overall accuracy (OA) increased from approximately 81% to 85.6%, indicating that the raster patching process successfully corrected most localized classification errors without altering the overall land use pattern. These findings suggest that most of the errors in the initial classification were not caused by the failure of the Random Forest model itself, but rather by spectral similarity among certain land use classes.

The land use classification achieved an OA of 85.6% and a Kappa coefficient of 0.82. As shown in Table 3, most land use classes were successfully identified, although several misclassifications were still observed among certain classes. The most prominent misclassification occurred between the settlement and water body classes, as well as between forest and dryland agriculture. These errors were likely caused by the spectral similarity between land cover class, the presence of mixed pixels along class boundaries, and the spatial resolution limitations of the Sentinel-2 imagery used in this study. Nevertheless, the number of misclassified pixels was relatively small compared with the number of correctly classified pixels and therefore did not have a significant impact on the overall quality of the land use map (Figure 11).

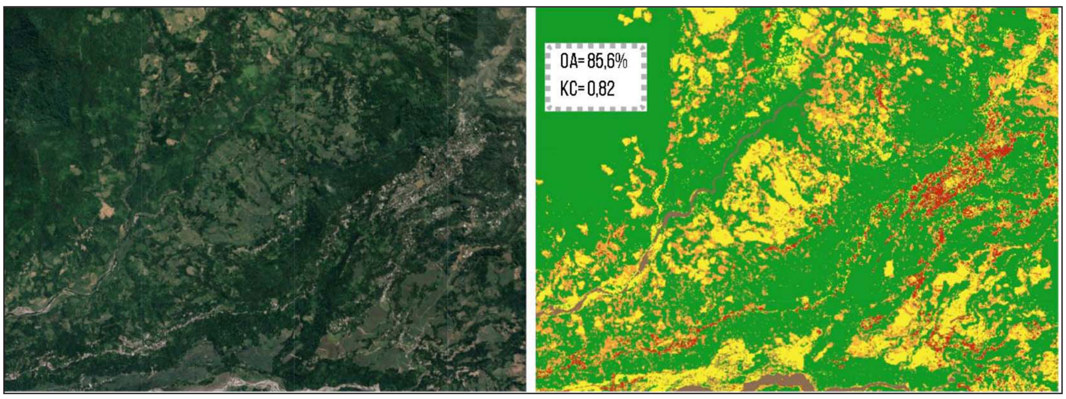

Figure 10. Accuracy improvement through raster patching

Table 3. Confusion matrix of the landuse classification

Land use	Land use prediction				
	Settlement	Water bodies	Wet land agriculuture	Dry land agriculuture	Forest
Settlement	1028	98	77	53	24
Water bodies	51	813	26	20	46
Wet land agriculuture	14	51	1022	27	42
Dry land agriculuture	13	11	23	920	78
Forest	11	14	44	91	1068
Metric				Value	Percent
Overall accuracy (OA)				0.856	85.6%
Kappa coefficient (KC)				0.82	

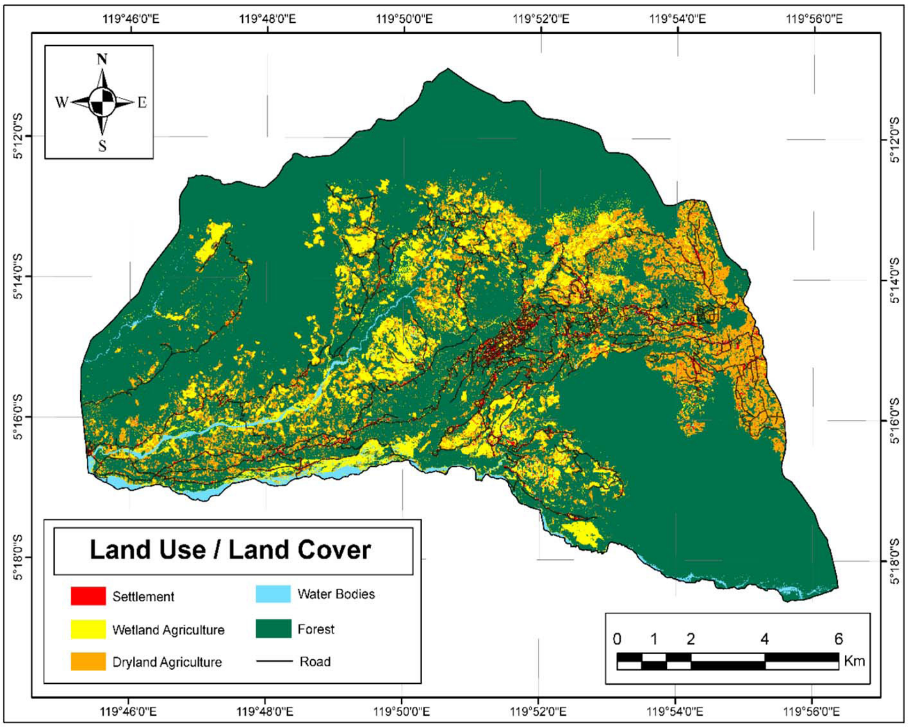

Figure 11. Landuse map

Land-use exposur

The analysis of land-use exposure serves as an advanced geospatial instrument to evaluate the complex interactions between physical susceptibility probabilities and the exposure of anthropogenic elements on the earth's surface. Rather than merely modeling slope instability zones in isolation, this stage integrates the probability outputs of the Random Forest algorithm with actual land cover data through spatial intersection techniques. This spatial overlay approach precisely delineates the extent to which vital infrastructure, built-up areas, and cultivation zones directly intersect with high-susceptibility matrices. Through this exposure mapping, blind spots in spatial planning can be visually identified, providing crucial scientific justification for the formulation of targeted ecological engineering and structural mitigation policies.

To delineate the priority mitigation areas, regions with high landslide susceptibility, specifically those with susceptibility values greater than 0.7, were quantitatively extracted and designated as high-priority zones. This specific threshold selectively delineates the most critical areas requiring immediate attention and intervention to support land-use planning and landslide disaster mitigation efforts (Figures 12, 13). The map above illustrates the land use classes located within areas of high landslide susceptibility, indicating that further management or mitigation measures are required in these areas to minimize potential risks. The extent of land use located within the high landslide susceptibility zone is presented in Table 4.

Spatial extraction revealed that the road network was the most proportionally exposed infrastructure, with 16.97% (53.28 ha) of its total area located directly within the high-hazard zone. Settlements also faced substantial exposure, with 4.90% (16.78 ha) of the built-up area situated within the prioritized risk zone. In contrast, although agricultural land occupied the largest total area across the landscape, it exhibited the lowest relative exposure, with only 0.22% (60.93 ha) located within the high-hazard zone.

Based on the spatial exposure map presented above, a distinct polarization of landslide threats was identified, showing a spatial distribution specifically associated with anthropogenic features. The exposure maps for the road network and built-up areas clearly display linear concentrations of hotspots at critical nodes, extending across Parigi, Garassi, Gantarang Villages, and Malino Subdistrict. This visual evidence confirms the existence of severe land-use conflicts, where the expansion of transportation networks and residential areas has substantially modified slope geometry, altered the natural slope conditions, and increased exposure to landslide processes in these steep terrain areas (Duan et al., 2025; Pradhan et al., 2022).

Meanwhile, the exposure of agricultural land exhibited a more dispersed spatial pattern than that of roads and settlements. This pattern indicates that agricultural activities have expanded into areas with high landslide susceptibility, thereby increasing the potential risk of reduced agricultural land productivity. However, within the study area, only a small proportion of agricultural land was located in highly susceptible zones, while most agricultural land, particularly in Bullutana, Pattapang, Bontolerung, and Parigi Villages, was situated within relatively safe zones.

In response to these spatial findings, the implementation of comprehensive land management interventions is essential. Sustainable land management measures may include imposing a moratorium on new construction permits within high-risk zones, implementing structural geotechnical engineering measures such as retaining wall installation along vulnerable road-cut slopes (Haris, 2024; Nuradrina, 2024), and integrating deep-rooted vegetative conservation techniques along agricultural land boundaries to stabilize soil aggregates ecologically without reducing land productivity (Fata et al., 2025). According to MacAfee et al. (2024), mitigation approaches that rely exclusively on geological

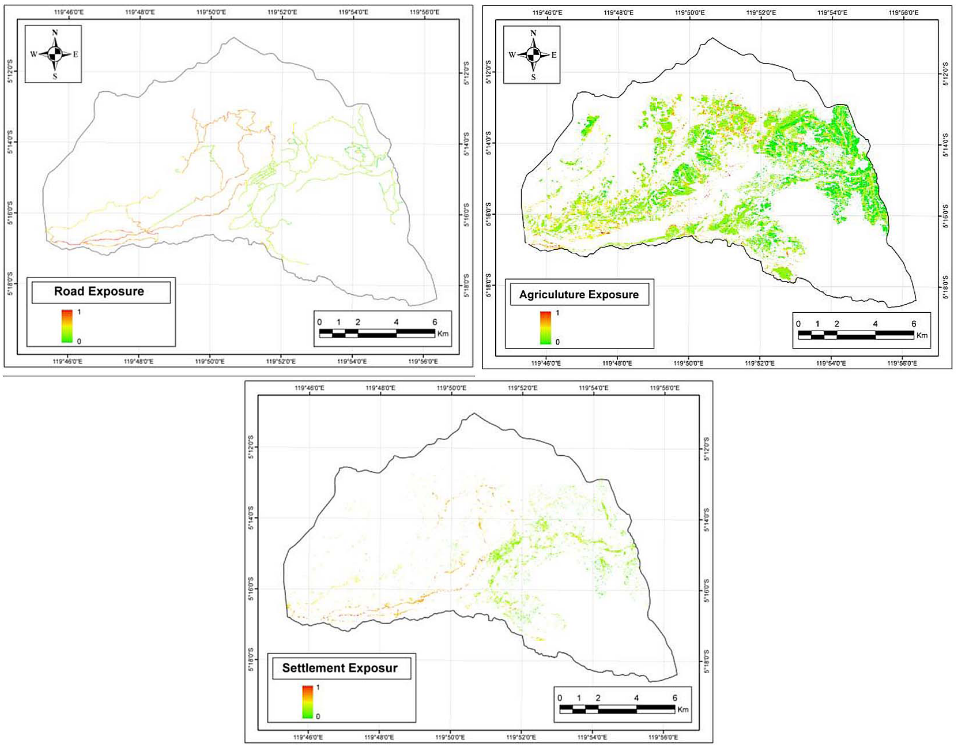

Figure 12. Land use exposure: (a) settlement (b) road (c) agriculture

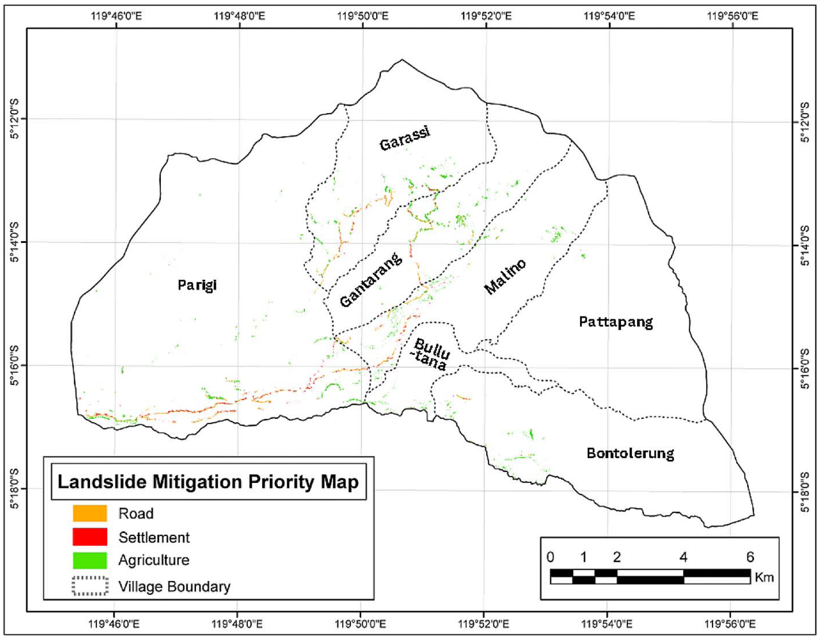

Figure 13. Landslide mitigation priority map

Table 4. Exposure of land use to high landslide susceptibility zones

Land use / Asset	Total area (ha)	Area exposed to high landslide suceptibily zone (ha)	Exposure (%)
Road	313.95	53.28	16.97
Settlement	342.16	16.78	4.9
Agricultural land	27,911.96	60.93	0.22

technical literature and physical infrastructure are often less effective because they overlook social processes at the local level. Therefore, mitigation efforts should also incorporate local knowledge to minimize landslide risk.

CONCLUSIONS

The results indicate that the integrated modeling framework achieved robust predictive reliability in landslide susceptibility mapping, as evidenced by an AUC value of 0.843, an Overall Accuracy of 80.3%, and consistent performance across classification metrics. The hypothesis that coupling Random Forest-based susceptibility modeling with high-resolution land-use classification significantly improves the interpretability and policy relevance of spatial risk assessments is fully supported. Rather than merely modeling slope instability in isolation, the integration successfully identified spatial hotspots where anthropogenic assets directly intersect with high-susceptibility matrices.

A key novel scientific contribution of this study is the explicit linkage of spatial hazard probabilities with a validated Sentinel-2 land-use classification (Overall Accuracy of 85.6%) within a single analytical framework. This addresses the critical gap in previous research that predominantly focused on susceptibility mapping as standalone products, without quantifying real-world anthropogenic exposure in rapidly developing mountainous regions.

The study further reveals that anthropogenic infrastructure, particularly roads, is disproportionately located within high susceptibility zones, with 16.97% of the total road network exposed to high-risk areas. Coupled with the finding that slope and distance to roads are the most dominant predictors of landslide occurrence, this confirms the critical role of human-induced slope modification in driving hazard distribution patterns. Additionally, the results show that improving predictive hazard accuracy alone is insufficient for spatial planning unless combined with this type of quantitative exposure assessment.

A remaining limitation of this research is that while it successfully identifies spatial exposure, it has not yet determined the specific mitigation priorities among the affected land uses and assets. Future prospects should focus on extending this framework by incorporating quantitative risk assessments, specifically by evaluating the monetary value of exposed infrastructure and assessing the direct hazard to human life. By integrating these economic and demographic dimensions, future studies can establish more comprehensive and highly targeted priority areas, ultimately guiding more efficient resource allocation for disaster mitigation.

Acknowledgements

The authors would like to express their sincere gratitude to Universitas Hasanuddin for providing academic support and research facilities throughout this study. We also thank the Regional Disaster Management Agency (BPBD) of Gowa Regency for providing data and supporting the field activities that greatly contributed to this research. Appreciation is also extended to the Environmental Management Study Program, Universitas Hasanuddin, for its support during the research process. Finally, the authors would like to thank all colleagues and individuals who assisted in data collection and provided valuable insights throughout the study.

REFERENCES

1. Akinci, H., Kilicoglu, C., Dogan, S. (2020). Random Forest-based landslide susceptibility mapping in coastal regions of Artvin, Turkey. *ISPRS International Journal of Geo-Information*, *9*(9), 553. https://doi.org/10.3390/ijgi9090553
2. Badan Nasional Penanggulangan Bencana. (2022, November 18). *Tanah longsor merusak jalan dan memakan korban jiwa di Gowa*. Badan Nasional Penanggulangan Bencana. https://bnpb.go.id/berita/tanah-longsor-merusak-jalan-dan-memakan-korban-jiwa-di-gowa
3. Badan Penanggulangan Bencana Daerah Provinsi Sulawesi Selatan. (2025). *Data bencana tanah longsor Kabupaten Gowa tahun 2025*. SIANDALAN. https://siandalan.sulselprov.go.id/
4. Baja, H., Kallenberg, M., Athanasiadis, I. N. (2025). To measure or not: A cost-sensitive, selective measuring environment for agricultural management decisions with reinforcement learning. *Proceedings of the AAAI Conference on Artificial Intelligence, 39*(27), 27831–27840.
5. Bayas, S., Sawant, S., Dhondge, I., Kankal, P., Joshi, A. (2022). Land use land cover classification using different ml algorithms on sentinel-2 imagery. In *Advanced machine intelligence and signal processing* (pp. 761–777). Singapore: Springer Nature Singapore.
6. Chen, N., Zang, B., Chen, Y. (2025). Predicting one-year post-surgical recurrence in colorectal liver metastasis using CT radiomics and machine learning. *PLoS One*, 20(8), e0330828. https://doi.org/10.1371/journal.pone.0330828
7. Chowdhury, M. S. (2023). A review on landslide susceptibility mapping research in Bangladesh. *Heliyon*, 9(7).
8. Çorbacıoğlu, Ş. K., Aksel, G. (2023). Receiver operating characteristic curve analysis in diagnostic accuracy studies: A guide to interpreting the area under the curve value. *Turkish Journal of Emergency Medicine*, 23(4), 195–198.
9. De Praga Baião, C. F., Mantovani, J., Alcântara, E. (2025). Increasing landslide susceptibility in urbanized areas of Petrópolis identified through spatio-temporal analysis. *Journal of South American Earth Sciences*, *160*, 105509. https://doi.org/10.1016/j.jsames.2025.105509
10. Dolchinkov, N. T. (2024). Natural Emergencies and Some Causes of Their Occurrence: a Review. *Trends in Ecological and Indoor Environmental Engineering, 2*(1), 18–27. https://doi.org/10.62622/TEIEE.024.2.1.18-27
11. Duan, Y., Ding, M., He, Y., Zheng, H., Delgado-Téllez, R., Sokratov, S.,..., Fuchs, S. (2025). Global projections of future landslide susceptibility under climate change. *Geoscience Frontiers*, 102074. https://doi.org/10.1016/j.gsf.2025.102074
12. Ejaz, F., Kausar, A., Ulfat, I., Vambol, S., Vambol, V., Lyndiuk, O. (2026). Geospatial assessment of urban environmental and mitigation strategies for thermal stress in Lahore, Karachi, and Islamabad. *Applied Geomatics*, *18*(3), 110. https://doi.org/10.1007/s12518-026-00759-2
13. Fata, Y. A., Hendrayanto, Erizal, Tarigan, S. D. (2025). Modeling of hydrological and mechanical effect of vegetation on landslide. *International Journal of Conservation Science*, *16*(2), 1013–1034.
14. Firoozi, A. A., Firoozi, A. A., Aati, K., Rashid, M. S. (2024). Integrated geotechnical modelling and real-time analysis for predicting earthquake-induced landslides and rockfalls in the East African fracture zone. *Trends in Ecological and Indoor Environmental Engineering, 2*(3), 1–19. https://doi.org/10.62622/TEIEE.024.2.3.01-19
15. Ganesh, B., Vincent, S., Pathan, S., Bhat, G. V. (2025). Optimized machine learning based model for the large-scale spatial prediction of landslides at western ghats in the state of Karnataka,

India. *Natural Hazards Research*.

16. Haris, A. D. (2024). *Analisa keruntuhan lereng pada jalan Muaradua-Kotabatu Kabupaten OKU Selatan* [Doctoral dissertation, Universitas Bina Darma].
17. Hussein, F., Latifi, H., Mojaradi, B. (2025). Leveraging Sentinel-2 A time series and integrated cellular automata-deep learning for accurate land cover forecasts in Northern Iraq. *Environmental earth sciences*, *84*(20), 586.
18. Liu, J., Wu, Y., Zhu, A., Ji, Z. (2025). Data-driven modeling and explainable optimization of organic micropollutant removal in forward osmosis membranes. *Journal of Environmental Chemical Engineering*, 120805.
19. MacAfee, E., Lohr, A. J., de Jong, E. (2024). Leveraging local knowledge for landslide disaster risk reduction in an urban informal settlement in Manado, Indonesia. *International Journal of Disaster Risk Reduction*, *111*, 104710. https://doi.org/10.1016/j.ijdrr.2024.104710
20. Merghadi, A., Yunus, A. P., Dou, J., Whiteley, J., ThaiPham, B., Bui, D. T., 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
21. Millimono, T. S. (2026). Risk production and urban vulnerability in Conakry: Integrating governance, land use, and biophysical factors in the 2025 Manéah Landslide. *Trends in Ecological and Indoor Environmental Engineering, 4*(2), 44–59. https://doi.org/10.62622/TEIEE.026.4.2.44-59
22. Nuradrina, A. B. (2024). *Analysis of slope stability with soil nailing reinforcement using PLAXIS program (Study case: STA 7+300 of Tawang-Ngalang segment IV road and bridge project)* [Doctoral dissertation, Universitas Islam Indonesia].
23. Phiri, D., Simwanda, M., Salekin, S., Nyirenda, V. R., Murayama, Y., Ranagalage, M. (2020). Sentinel-2 data for land cover/use mapping: A review. *Remote sensing*, *12*(14), 2291. https://doi.org/10.3390/rs12142291
24. Pradhan, S., Toll, D. G., Rosser, N. J., Brain, M. J. (2022). An investigation of the combined effect of rainfall and road cut on landsliding. *Engineering Geology*, *307*, 106787. https://doi.org/10.1016/j.enggeo.2022.106787
25. Rabby, Y. W., Li, Y., Abedin, J., Sabrina, S. (2022). Impact of land use/land cover change on landslide susceptibility in Rangamati Municipality of Rangamati District, Bangladesh. *ISPRS International journal of geo-information*, *11*(2), 89. https://doi.org/10.3390/ijgi11020089
26. Tsuchiya, S., Koga, S., Sasahara, K., Matsui, M., Nakahiro, M., Watanabe, H., Shima., Yoshida, K. (2004). Reconnaissance of the gigantic landslide occurred on Mt. Bawakaraeng in the south Sulawesi state of Indonesia and unstable debris sedimentation (prompt report). *Journal of the Japan Society of Erosion Control Engineering*, *57*(3), 40–46.
27. Vodde, F., Ait, K., Orumaa, A., Jogiste, K., Kaart, T., Kiviste, A.,..., Metslaid, M. (2025). Drivers behind the spatial dispersion of European spruce bark beetle (Ips typographus) infestation in protected areas in Estonia, four years after a major storm. *Forest Ecology and Management*, *578*, 122469.
28. Wang, M., Zhao, G., Wang, S. (2024). Hybrid random forest models optimized by Sparrow search algorithm (SSA) and Harris hawk optimization algorithm (HHO) for slope stability prediction. *Transportation Geotechnics*, *48*, 101305.
29. Wu, M., Zhang, G., Zhang, S., Huang, W., Zou, H. (2025). Spatiotemporal decoupling of rainfall patterns and shallow landslide stability: Lessons learned from calibrated ensemble physically-based models. *Journal of Rock Mechanics and Geotechnical Engineering*.
30. Xing, Y., Li, J., Wang, P. (2026). Estimating landslide triggering rainfall thresholds: a probabilistic framework for susceptibility analysis. Advances in Space Research.
31. Yao, R., Bekhor, S. (2020). Data-driven choice set generation and estimation of route choice models. *Transportation Research Part C: Emerging Technologies*, *121*, 102832. https://doi.org/10.1016/j.trc.2020.102832
32. Yulianto, F., Nugroho, G., Chulafak, G. A., Suwarsono, S. (2021). Improvement in the accuracy of the postclassification of land use and land cover using Landsat 8 data based on the majority of segment-based filtering approach. *The Scientific World Journal*, 2021, 6658818. https://doi.org/10.1155/2021/6658818
33. Zandi, R., Far, G. S. P. (2025). Evaluating the factors affecting landslides using machine learning algorithms (case study: the catchment area of Karun-3 Dam, Iran). *The Egyptian Journal of Remote Sensing and Space Sciences*, *28*(3), 512–522.
34. Zhang, Q., He, Y., Zhang, L., Lu, J., Gao, B., Yang, W.,..., Zhang, Y. (2024). A landslide susceptibility assessment method considering the similarity of geographic environments based on graph neural network. *Gondwana Research*, *132*, 323–342. https://doi.org/10.1016/j.gr.2024.04.013