Equivalent: Jurnal Ilmiah Sosial Teknik
Volume 8, Issue 5, 2111-2130
e_ISSN: 2775-0833
https://jurnalequivalent.id/index.php/jequi/index
DOI: doi.org/10.59261/jequi.v8i5.459

Interpretable Landslide Susceptibility Mapping in Pacitan, Indonesia Using Bayesian-Optimized XGBoost and SHAP Analysis

Aghasi Yanda Wafa Azizah
Universitas Negeri Surabaya, Indonesia

Khalila Dzakia Agram
Universitas Negeri Surabaya, Indonesia

Shellomitha Sulvana Dewi
Universitas Negeri Surabaya, Indonesia

Saira Nashwa Ramadhani
Universitas Negeri Surabaya, Indonesia

Yuliani Puji Astuti*
Universitas Negeri Surabaya, Indonesia

***Corresponding author:**
Yuliani Puji Astuti, Universitas Negeri Surabaya, Indonesia.
✉yulianipuji@unesa.ac.id

Article Info:

Article history:
Received: September 07, 2026
Revised: October 10, 2026
Accepted: October 12, 2026
Available Online: November 04, 2026

Keywords:
Bayesian Optimization;
Landslide Susceptibility Mapping;
QGIS;
Pacitan;
SHAP;
XGBoost;

This is an open access article under the CC BY-SA license.

Copyright © 2026 by Author. Published by Cerdika International Cyber Polytechnic

Abstract

Background: Landslide susceptibility mapping is essential in areas with complex topographic and geological conditions, such as Pacitan Regency, Indonesia, to identify landslide-prone areas and assess the factors influencing landslide susceptibility.
Objective: This study developed a landslide susceptibility mapping model using Extreme Gradient Boosting (XGBoost), optimized through Bayesian optimization and interpreted using the SHapley Additive exPlanations (SHAP) method.
Methods: The dataset consisted of 1,486 samples, comprising 743 landslide occurrence points and 743 non-landslide points. Eleven landslide conditioning factors were used as predictor variables: elevation, aspect, slope, geology, land use, profile curvature, plan curvature, distance to roads, distance to rivers, Stream Power Index (SPI), and Topographic Wetness Index (TWI). Model hyperparameter optimization was performed using Bayesian optimization with a Tree-structured Parzen Estimator (TPE).
Results: The optimized XGBoost model achieved an accuracy of 69.80%, a precision of 70.34%, a recall of 68.46%, an F1 score of 69.39%, and a receiver operating characteristic area under the curve (ROC-AUC) of 0.7564 on the independent test set. The resulting susceptibility map classified the study area into five categories: very low, low, moderate, high, and very high susceptibility. SHAP analysis identified slope, aspect, plan curvature, geology, and distance to roads as the dominant factors influencing model predictions.
Conclusion: Integrating Bayesian optimization, XGBoost, and SHAP provides an interpretable approach to landslide susceptibility mapping in Pacitan Regency. The resulting susceptibility map and feature-level explanations can support preliminary disaster risk assessment and spatial planning. However, residual classification errors and limitations associated with the use of static environmental predictors should be considered when interpreting the results.

To cite this article: Azizah, A. Y. W., Agram, K. D., Dewi, S. S., Ramadhan, S. N., & Astuti, Y. P. (2026). Interpretable landslide susceptibility mapping in Pacitan, Indonesia using Bayesian-optimized XGBoost and SHAP analysis. *Equivalent: Jurnal Ilmiah Sosial Teknik*, *8*(5), 2111-2130. https://doi.org/10.59261/jequi.v8i5.459

INTRODUCTION

Landslides are significant geological hazards that threaten communities, infrastructure, ecosystems, and economic activities, particularly in mountainous regions such as Pacitan Regency, East

Java, where Tropical Cyclone Cempaka triggered slope failures in November 2017 (Samodra, 2024). Landslide occurrence is influenced by complex interactions among topography, geology, hydrology, land use, and human activities, resulting in nonlinear relationships and spatial variability (Ado et al., 2022; Hussain et al., 2022). Consequently, landslide susceptibility mapping (LSM) is essential for identifying areas susceptible to landslides and supporting spatial planning, infrastructure development, and disaster risk mitigation.

Traditional LSM approaches, including the Analytical Hierarchy Process (AHP), frequency ratio, and logistic regression, offer interpretable relationships between conditioning factors and landslide occurrence but have limitations in capturing nonlinear interactions (Ado et al., 2022). Advances in remote sensing, geographic information systems (GIS), and the availability of spatial data have facilitated the application of machine learning techniques. Among these, Extreme Gradient Boosting (XGBoost) is an ensemble learning algorithm that sequentially constructs decision trees to minimize prediction errors, enabling effective modeling of complex relationships (Chen & Guestrin, 2016). However, its predictive performance depends heavily on hyperparameter configurations, including the learning rate, tree depth, number of estimators, subsampling rate, and regularization parameters.

Hyperparameter optimization is therefore essential for improving model performance and generalization (Bischl et al., 2021). Although grid search systematically evaluates parameter combinations, its computational requirements can increase substantially with the size and dimensionality of the search space, whereas random search does not explicitly use the results of previous evaluations to guide subsequent trials (Bergstra & Bengio, 2012). Bayesian optimization addresses these limitations by using a probabilistic surrogate model and an acquisition function to identify promising hyperparameter configurations adaptively (Snoek et al., 2012). This study employs Bayesian optimization using the tree-structured Parzen estimator (TPE) to optimize XGBoost hyperparameters efficiently (Bergstra et al., 2011).

Beyond predictive accuracy, model interpretability remains a critical challenge because machine learning algorithms frequently operate as black-box systems (Guidotti et al., 2018). In landslide susceptibility assessment, understanding the contributions of environmental factors is essential for interpreting predictions scientifically and supporting transparent decision-making. SHapley Additive exPlanations (SHAP), which is grounded in cooperative game theory, addresses this challenge by quantifying individual feature contributions at both the global and local levels (Lundberg & Lee, 2017). Unlike conventional feature importance measures, SHAP also indicates the direction and magnitude of each feature's contribution to model predictions (Lundberg et al., 2020).

This study incorporates eleven landslide conditioning factors: elevation, aspect, slope, geology, land use, profile curvature, plan curvature, distance to roads, distance to rivers, Stream Power Index (SPI), and Topographic Wetness Index (TWI). These variables represent geomorphological, geological, hydrological, land-cover, and anthropogenic conditions associated with slope instability (Hussain et al., 2022). Morphometric variables describe terrain characteristics, geology represents the properties of slope materials, land use and distance to roads reflect land-cover patterns and human influences, and distance to rivers, SPI, and TWI characterize aspects of hydrological processes, including stream power and topographic wetness.

Pacitan Regency is a relevant study area because of its mountainous terrain and diverse geological and land-use characteristics (Putra et al., 2021). Previous studies have demonstrated the applicability of GIS and machine learning techniques for mapping landslide susceptibility in this region using environmental conditioning factors (Darminto et al., 2021; Dimyati et al., 2022). However, some studies have primarily emphasized predictive accuracy without simultaneously integrating hyperparameter optimization and model interpretation. Although Bayesian optimization and SHAP have previously been combined in landslide susceptibility research Sun et al. (2023), further investigation is warranted to integrate predictive performance and interpretability within the specific geomorphological and disaster context of Pacitan Regency.

To address this gap, this study integrates XGBoost, Bayesian optimization, and SHAP to model landslide susceptibility using 1,486 balanced sample points (743 landslide and 743 non-landslide locations) and eleven raster-based conditioning factors. Its novelty lies in applying this framework to a field-verified inventory of cyclone-triggered landslides in Pacitan Regency, optimizing hyperparameters

through training cross-validation, and interpreting feature contributions in relation to local topographic conditions. The study aims to develop and evaluate the model, identify feature importance and directional effects using SHAP, and generate a probability-based susceptibility map to support transparent disaster risk assessment and spatial decision-making (Lundberg et al., 2020).(Lundberg et al., 2020).

METHOD

This research was conducted in Pacitan Regency, East Java Province, Indonesia. Pacitan Regency is in the southwestern part of East Java Province and has diverse physiographic characteristics, ranging from coastal areas to hilly and mountainous terrain. These geomorphological conditions result in variations in topography, slope, elevation, surface morphology, and surface runoff patterns, all of which influence slope stability (Darminto et al., 2021).

Administratively, Pacitan Regency is bordered by Wonogiri Regency in Central Java Province to the north and west, Ponorogo and Trenggalek Regencies to the east, and the Indian Ocean to the south. The predominance of hilly and mountainous terrain, together with diverse geological conditions and land-use patterns, makes Pacitan Regency an area susceptible to landslides.



Figure 1. Study Area Pacitan



Figure 2. Landslide Cases

Variations in topographic, geological, hydrological, and land-use conditions contribute to spatial variations in landslide susceptibility across Pacitan Regency. Therefore, this area was selected as the study area to develop a machine learning-based model for identifying relationships between environmental conditioning factors and landslide occurrences.

In this study, spatial data covering Pacitan Regency were used as the basis for pixel-based landslide susceptibility modeling. Landslide conditioning factors served as predictor variables, whereas landslide and non-landslide inventory data were used to train and test the model. The best-performing model obtained through hyperparameter optimization was subsequently applied to the entire study area to produce a landslide susceptibility map.

This study used secondary data obtained from Mendeley Data covering Pacitan Regency, East Java, Indonesia (Samodra, 2024). The dataset provided spatial data and a landslide inventory for developing a machine learning-based landslide susceptibility mapping model. It contained information on landslide and non-landslide locations, as well as various environmental factors required to analyze the relationships between regional characteristics and the spatial distribution

of landslide occurrences (Samodra, 2024). The landslide inventory was verified through field surveys, and the conditioning factors were derived from topographic maps and the National Digital Elevation Model (DEMNAS) (Samodra, 2024).

The dataset consisted of a landslide and non-landslide inventory and raster data representing landslide conditioning factors. The inventory data were used to identify sample locations representing two target classes: landslide locations (class 1) and non-landslide locations (class 0). The conditioning-factor rasters were used to obtain the values of the predictor variables by extracting raster values at the sample locations.

The processed dataset comprised 1,486 sample points, consisting of 743 landslide locations and 743 non-landslide locations. The balanced sample distribution was intended to reduce potential class imbalance during model training (Krawczyk, 2016). However, the equal proportions of the two classes reflected the sampling design of the benchmark dataset rather than the natural prevalence of landslides in Pacitan Regency (Samodra, 2024). Each sample point was associated with eleven conditioning factors extracted from raster data to construct the final classification dataset. Potential uncertainty associated with non-landslide labels was considered when interpreting the model's predictive performance.

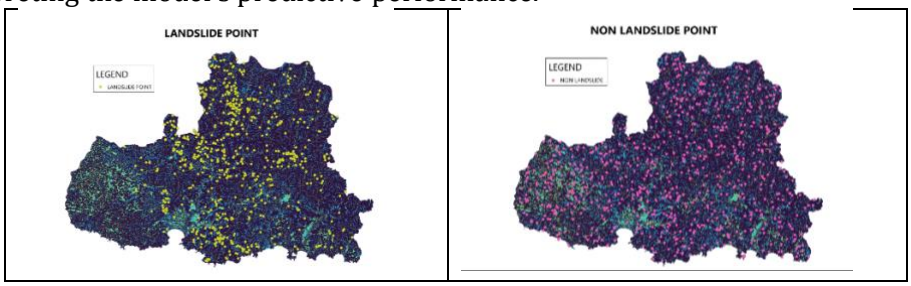
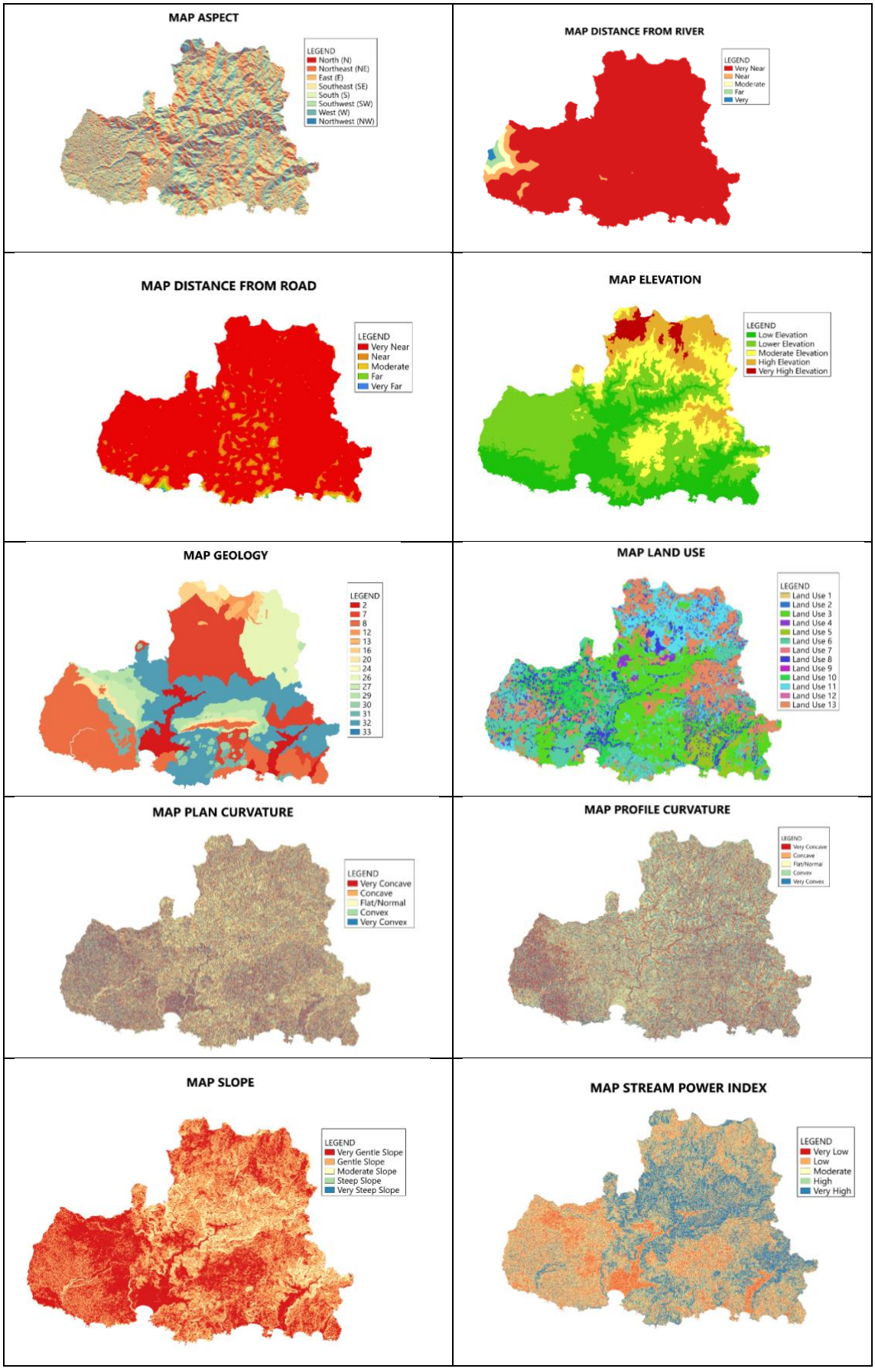


Figure 3. Landslide Point and Non Landslide Point

A total of eleven landslide conditioning factors were used as predictor variables: elevation, aspect, slope, geology, land use, profile curvature, plan curvature, distance to roads, distance to rivers, Stream Power Index (SPI), and Topographic Wetness Index (TWI). These factors were selected because they represent topographic characteristics, geological conditions, hydrological processes, and anthropogenic influences that may affect the spatial distribution of landslide occurrences (Hussain et al., 2022).

Initial spatial data processing was performed using QGIS to ensure consistency in coordinate reference systems, spatial coverage, and raster-layer alignment. This stage included checking data quality, extracting raster values at sample locations, and compiling the final dataset in comma-separated values (CSV) format. The processed dataset was then used as input for modeling in Python through Google Colaboratory (Google Colab). The modeling workflow included data preprocessing, dataset splitting, XGBoost hyperparameter optimization using Bayesian optimization, model performance evaluation, and interpretability analysis using SHAP. XGBoost is a gradient-boosting-based ensemble learning algorithm capable of modeling complex relationships in classification data Chen & Guestrin (2016), whereas Bayesian optimization efficiently searches for promising hyperparameter configurations using information from previous evaluations (Snoek et al., 2012). SHAP analysis was applied to explain the contributions of individual predictor variables to model predictions based on the Shapley value concept (Lundberg & Lee, 2017).

A total of eleven landslide conditioning factors were used as predictor variables in this study. These factors were selected to represent topographic, geological, land-use, and hydrological characteristics, as well as conditions associated with human activities that may contribute to landslide susceptibility.


MAP ASPECT
LEGEND
North (N)
Northeast (NE)
East (E)
Southeast (SE)
South (S)
Southwest (SW)
West (W)
Northwest (NW)
MAP DISTANCE FROM RIVER
LEGEND
Very Near
Near
Moderate
Far
Very
MAP DISTANCE FROM ROAD
LEGEND
Very Near
Near
Moderate
Far
Very Far
MAP ELEVATION
LEGEND
Low Elevation
Lower Elevation
Moderate Elevation
High Elevation
Very High Elevation
MAP GEOLOGY
LEGEND
2
7
8
12
13
16
20
24
26
27
29
30
31
32
33
MAP LAND USE
LEGEND
Land Use 1
Land Use 2
Land Use 3
Land Use 4
Land Use 5
Land Use 6
Land Use 7
Land Use 8
Land Use 9
Land Use 10
Land Use 11
Land Use 12
Land Use 13
MAP PLAN CURVATURE
LEGEND
Very Concave
Concave
Flat/Normal
Convex
Very Convex
MAP PROFILE CURVATURE
LEGEND
Very Concave
Concave
Flat/Normal
Convex
Very Convex
MAP SLOPE
LEGEND
Very Gentle Slope
Gentle Slope
Moderate Slope
Steep Slope
Very Steep Slope
MAP STREAM POWER INDEX
LEGEND
Very Low
Low
Moderate
High
Very High

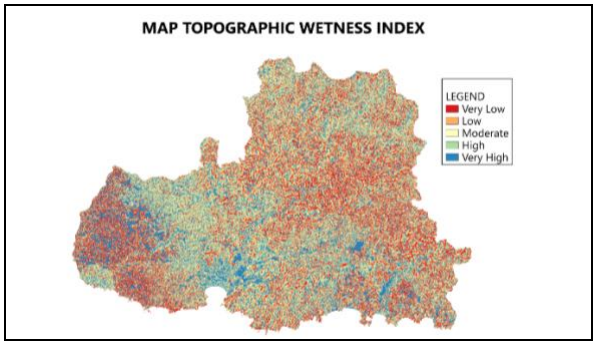


Figure 4. Conditioning factors

The factors used consist of:

Table 1

Conditioning Factors

No	Conditioning Factors	Abbreviations/Descriptions
1	Elevation	Elevation
2	Aspect	Slope direction/orientation
3	Slope	Slope gradient
4	Geology	Geological conditions
5	Land Use	Land use
6	Profile Curvature	Curvature profile
7	Plan Curvature	Curvature of the plan
8	Distance to Road	Distance to road
9	Distance to River	Distance to the river
10	Stream Power Index	SPI
11	Topographic Wetness Index	TWI

The eleven factors were used consistently throughout the modeling and mapping processes. The topographic factors comprised elevation, aspect, slope, profile curvature, and plan curvature. Geological and land-use factors represented geological characteristics and land-surface use, respectively. Distance to roads and distance to rivers represented the spatial relationships between sample locations and road networks and river systems. Meanwhile, SPI and TWI represented characteristics associated with surface-flow processes and topographic wetness conditions.

The preprocessing stage was conducted to prepare the spatial and tabular data for use in machine learning models. This stage was important because the quality of the input data affects a model's ability to identify patterns and generate reliable predictions (Kotsiantis et al., 2006). Initial spatial data processing was performed using QGIS, whereas tabular data preprocessing was conducted using Python in Google Colab.

First, all conditioning-factor rasters were checked to ensure that they had compatible coordinate reference systems and consistent spatial coverage. Landslide and non-landslide inventory points were then used as sample locations for extracting values from each conditioning-factor raster. Combining landslide inventory data with environmental conditioning factors is a common approach in landslide susceptibility modeling for establishing relationships between landslide locations and environmental characteristics (Ado et al., 2022).

Raster values were extracted using the Sample Raster Values function in QGIS. Each sample point was assigned values from the eleven raster layers, linking each sample to the environmental characteristics of its location.

Following raster-value extraction, the dataset was checked for missing values, NoData values, duplicate records, and data-type inconsistencies. The target variable was then encoded as a binary variable, as follows:

$$y = \begin{cases} 1, & \text{Landslide} \\ 0, & \text{Non-Landslide} \end{cases}$$

Columns that did not serve as predictor or target variables, such as feature identifiers, coordinates, layer information, and file paths, were excluded from the modeling process.

The final dataset was organized into a feature matrix ((X)) containing the eleven landslide conditioning factors and a target vector ((y)) containing the corresponding binary class labels for landslide and non-landslide locations.

The preprocessed dataset was divided into training and testing subsets using an 80:20 stratified split to preserve the proportions of landslide and non-landslide samples. This procedure resulted in 1,188 training samples and 298 testing samples. The training subset was used for XGBoost model development and Bayesian hyperparameter optimization, whereas the testing subset was reserved for evaluating predictive performance on previously unseen observations from the same dataset. A random state of 42 was used to ensure reproducibility of the data partition. Model performance was evaluated using accuracy, precision, recall, F1-score, and the area under the receiver operating characteristic curve (ROC-AUC).

The Extreme Gradient Boosting (XGBoost) model was used to classify landslide and non-landslide locations based on the selected conditioning factors. XGBoost is a gradient-boosting-based ensemble learning algorithm that sequentially builds decision trees, with each successive tree contributing to the optimization of the model's objective function (Chen & Guestrin, 2016).

Mathematically, the final prediction of the XGBoost model can be expressed as:

$$\hat{y}_i = \sum_{k=1}^{K} f_k(x_i), \quad f_k \in F$$

by being the predictive value for observation to, indicating the number of $\hat{y}_i - i$Kdecision trees in the XGBoost model, indicating the function of the Decision tree showing the predictor variable vector representing the landslide conditioning factor in observation to.$f_k$ke − k, serta $x_i - i$

In this study, XGBoost was used for binary classification to distinguish landslide from non-landslide locations because of its ability to capture nonlinear relationships and complex interactions among predictor variables. The XGBoost hyperparameters were subsequently optimized using Bayesian optimization to identify the best configuration based on cross-validation performance on the training data.

Bayesian optimization was used to identify an optimal configuration of XGBoost hyperparameters. This method efficiently searches for promising hyperparameter combinations by incorporating information from previous evaluations through a sequential optimization procedure (Akiba et al., 2019). The optimization was performed using the Tree-structured Parzen Estimator (TPE) approach. TPE models the distributions of hyperparameter configurations associated with different objective-function outcomes and uses these distributions to propose subsequent candidates expected to yield improved model performance.

The optimization process was conducted using the training data and a cross-validation approach. Each hyperparameter configuration was evaluated using the mean cross-validation ROC-AUC as the objective function. ROC-AUC was selected because it measures the model's ability to discriminate between landslide and non-landslide classes across different classification thresholds.

The following hyperparameters were optimized.

Table 2
Hyperparameter

Hyperparameter	Function
n_estimators	Regulates the number of decision trees built in the boosting process. Higher values allow the model to learn more complex patterns, but can increase the risk of overfitting.
learning_rate	Control the contribution of each tree to the final model. Smaller values result in a

	more gradual and stable learning process.
max_depth	Specifies the maximum depth of each decision tree. This parameter regulates the level of complexity of the model in capturing data patterns.
min_child_weight	Sets the minimum amount of sample weight required to form a new node on the tree. This parameter helps control complexity and prevents the formation of trees that are too specific to the training data.
Subsample	Determine the proportion of training samples used to build each tree. These parameters are used to improve the model's generalization capabilities and reduce the risk of overfitting.
colsample_bytree	Sets the proportion of features used in each tree formation. These parameters help reduce the model's dependence on certain variables.
Gamma	Determine the minimum reduction of the loss function required for a split in the tree to be performed. Higher values make the model more conservative.
reg_alpha	Control the L1 regularization of the model to reduce complexity by penalizing large feature weights.
reg_lambda	Control the L2 regularization of the model to limit model complexity and improve prediction stability.

The nine hyperparameters were used to define the search space for Bayesian optimization. Each configuration proposed by the optimization algorithm was evaluated using the mean cross-validation ROC-AUC as the objective function. The configuration that achieved the highest mean cross-validation ROC-AUC was selected as the best configuration and used to train the final XGBoost model.

The performance of the final XGBoost model was evaluated using the testing data, which had not been used during model training or hyperparameter optimization. The evaluation assessed the model's ability to classify landslide and non-landslide locations using accuracy, precision, recall, F1-score, and ROC-AUC. In addition, a confusion matrix was used to describe the distribution of prediction outcomes in terms of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) (Grandini et al., 2020).

Accuracy represents the proportion of correctly classified observations among all test observations. It is calculated as follows:

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

Precision represents the proportion of correctly identified landslide locations among all locations predicted to be landslides. It is calculated as follows:

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

Recall, or sensitivity, represents the proportion of actual landslide locations correctly identified by the model. It is calculated as follows:

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

The F1-score is the harmonic mean of precision and recall. It provides a balance between identifying landslide locations and limiting false-positive and false-negative classifications. It is calculated as follows:

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

In addition to these metrics, the area under the receiver operating characteristic curve (ROC-AUC) was used to evaluate the model's ability to discriminate between landslide and non-landslide classes across different probability thresholds. Higher ROC-AUC values indicate better discrimination between the two classes.

After the XGBoost model with the best hyperparameter configuration had been obtained, SHapley Additive exPlanations (SHAP) analysis was performed to interpret the contributions of individual landslide conditioning factors to the model's predictions. SHAP was used because machine learning models such as XGBoost have complex structures and nonlinear relationships among variables that cannot be fully explained by built-in feature importance measures alone (Khan Inan & Rahman, 2022; Ponce-Bobadilla et al., 2024). SHAP provides model explanations by quantifying feature contributions based on Shapley values, both globally and for individual observations, thereby helping explain how different factors contribute to model predictions (Ponce-Bobadilla et al., 2024).

SHAP assigns a contribution value to each feature for a given prediction based on the concept of Shapley values from cooperative game theory. The model output can generally be expressed as:

$$f(x) = \phi_0 + \sum_{j=1}^{M} \phi_j$$

By being the model's predictive value, it is the base f(x) $\phi_0$value of the model, it is the number of predictor variables, and it is the contribution of conditioning factors to the prediction results. $\phi_0$ke − j

SHAP analysis was conducted to identify the conditioning factors that contributed most strongly to landslide susceptibility predictions and to examine the direction of each feature's contribution to the model output. The global importance of each feature was quantified using its mean absolute SHAP value. Higher mean absolute SHAP values indicate greater average contributions to the model's predictions (Lundberg et al., 2020).

The SHAP summary plot was used to visualize the global importance of the conditioning factors and the distribution of their contributions to the model output, including the direction of the relationship between feature values and model predictions (Lundberg et al., 2020). In addition, SHAP dependence plots were used to examine the relationships between individual conditioning-factor values and their corresponding SHAP values, providing insight into how changes in feature values were associated with changes in the model output (Molnar, 2020).

Landslide susceptibility mapping was conducted after the XGBoost model with the best hyperparameter configuration had been obtained through Bayesian optimization. The model was then applied to the entire study area using the landslide conditioning-factor rasters as spatial predictor variables (Ado et al., 2022).

A total of eleven landslide conditioning-factor rasters were used as model inputs to generate predictions for each pixel in the study area. The raster values were arranged in the same order as the predictor variables used during model training: elevation, aspect, slope, geology, land use, profile curvature, plan curvature, distance to roads, distance to rivers, SPI, and TWI.

The predictor variables were arranged in this order to ensure consistency between the spatial data structure used during mapping and the feature vectors used to train the XGBoost model. The model then generated a predicted probability of landslide occurrence for each pixel based on the combination of conditioning-factor values at that location.

The predicted probability was used as the landslide susceptibility value. Mathematically, the susceptibility value for each pixel can be expressed as:

$$S(x,y) = P(Y = 1|X_{x,y})$$

With:

$S(x,y)$ is the value of landslide susceptibility on pixels with coordinates, while it is the probability of landslides based on the value of the conditioning factor at the location of the pixel.$(x,y)P(Y = 1|X_{x,y})X$

The susceptibility values ranged from 0 to 1. Values closer to 1 indicated a higher model-predicted probability of landslide occurrence, whereas values closer to 0 indicated a lower predicted probability. The probability predictions were converted into a landslide susceptibility raster and visualized using QGIS.

The susceptibility raster was subsequently classified into five susceptibility levels using the quantile classification method: Very Low, Low, Moderate, High, and Very High. The quantile method divided the distribution of predicted susceptibility values into classes containing approximately equal numbers of pixels, subject to the distribution of the data and the classification implementation. The resulting classification was used to produce a landslide susceptibility map of Pacitan Regency showing the spatial distribution of susceptibility levels predicted by the XGBoost model.

The research workflow encompassed data collection, spatial data processing, classification model development, model interpretation, and the production of a landslide susceptibility map of Pacitan Regency. The main stages consisted of dataset preparation, data preprocessing, XGBoost modeling, hyperparameter optimization using Bayesian optimization, model performance evaluation, interpretability analysis using SHAP, and landslide susceptibility mapping.

Overall, the research workflow consisted of the following stages:

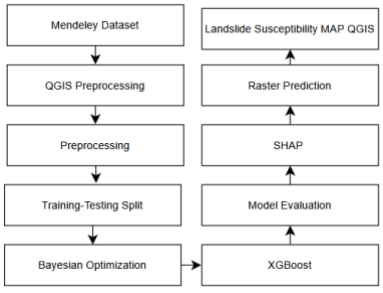


Figure 5. Research flow

RESULTS AND DISCUSSION

Results

Dataset Characteristics

The spatial data processing results yielded the final dataset used for landslide susceptibility modeling in Pacitan Regency. The dataset consisted of 1,486 sample points, including 743 landslide points and 743 non-landslide points. Landslide locations were labeled as class 1, whereas non-landslide locations were labeled as class 0. These labels served as the target variable in the classification process.

Each sample point was associated with values for eleven landslide conditioning factors, which served as predictor variables in the XGBoost model. These factors included elevation, aspect, slope, geology, land use, profile curvature, plan curvature, distance to road, distance to river, Stream Power Index (SPI), and Topographic Wetness Index (TWI).

The distribution of samples across the target classes is presented in Table 3.

Table 3

Dataset characteristics

Classes	Label	Number of Samples	Percentage
Landslide	1	743	50%

Non Landslide	0	743	50%
Total	-	1.486	100%

As shown in Table 3, the dataset exhibited a balanced class distribution, with equal numbers of landslide and non-landslide samples. This balance was intended to reduce potential bias toward either class during model training, allowing the model to learn patterns associated with both classes.

Furthermore, the dataset was divided into training and testing subsets using a stratified split with an 80:20 ratio to maintain the class distribution in each subset. This procedure yielded 1,188 training samples (80%) and 298 testing samples (20%), as presented in Table 4. The training data were used to train the XGBoost model and optimize its hyperparameters through Bayesian optimization, whereas the testing data were reserved for evaluating model performance.

Table 4
Dataset

Dataset	Number of Samples	Percentage
Training	1.188	80%
Testing	298	20%
Total	1.486	100%

Bayesian Optimization Results

Bayesian optimization was performed to identify an optimal configuration of XGBoost hyperparameters based on a predefined objective function. The optimization process employed the Tree-structured Parzen Estimator (TPE) approach, in which each hyperparameter configuration was evaluated through a series of trials. TPE builds a probabilistic model from the history of previous evaluations to select subsequent configurations that are likely to improve model performance efficiently (Dasgupta & Sen, 2024; Watanabe, 2023).

The optimization process identified a hyperparameter configuration that achieved a mean cross-validation area under the receiver operating characteristic curve (ROC-AUC) of 0.7541. This value indicates the model's ability to discriminate between the landslide and non-landslide classes during cross-validation. ROC-AUC was selected as an evaluation metric because it measures a model's ability to discriminate between two classes across different classification thresholds (Grandini et al., 2020). The optimal hyperparameter configuration identified through Bayesian optimization is presented in Table 5.

Table 5
Hyperparameter XGBoost results of Bayesian Optimization

Hyperparameter	Best value
n_estimators	850
learning_rate	0,0307
max_depth	7
min_child_weight	3
Subsample	0,7926
colsample_bytree	0,8357
Gamma	3,9590
reg_alpha	0,00685
reg_lambda	8.16×10^{-6}

The optimal hyperparameter configuration was subsequently used to construct the final XGBoost model. These results indicate that the model was configured through a Bayesian optimization-based search rather than being trained with the default hyperparameter settings.

This approach aimed to improve predictive performance in capturing nonlinear relationships between environmental conditioning factors and landslide occurrence.

Performa Model XGBoost

The final XGBoost model was constructed using the optimal hyperparameter configuration identified through Bayesian optimization. Model performance was evaluated using the 298 testing samples, which had not been used during model training or hyperparameter optimization. The model was evaluated using several classification metrics: accuracy, precision, recall, F1-score, and ROC-AUC. The evaluation results for the testing dataset are presented in Table 6.

Table 6
Performance model XGBoost

Metrics	Value
Accuracy	0,6980
Precision	0,7034
Recall	0,6846
F1 Score	0,6939
ROC-AUC	0,7564

The accuracy value indicates that the model correctly classified approximately 69.80% of the testing samples. Precision represents the proportion of correctly predicted landslide samples among all samples predicted as landslides. Recall represents the proportion of actual landslide samples correctly identified by the model.

The F1-score of 0.6939 summarizes the balance between precision and recall. Meanwhile, the ROC-AUC value of 0.7564 indicates the model's ability to discriminate between landslide and non-landslide classes based on its predicted scores across different classification thresholds.

Confusion Matrix

In addition to the evaluation metrics, a confusion matrix was used to examine the distribution of correct and incorrect classifications produced by the XGBoost model on the testing dataset.

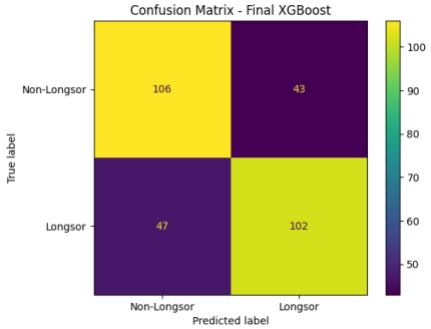


Figure 6
Confusion Matrix Model XGBoost

The confusion matrix comprises four main components:

1. True positive (TP): The number of landslide locations correctly classified as landslides.
2. True negative (TN): The number of non-landslide locations correctly classified as non-landslides.
3. False positive (FP): The number of non-landslide locations incorrectly classified as landslides.
4. False negative (FN): The number of landslide locations incorrectly classified as non-landslides.

Based on the confusion matrix, the XGBoost model correctly classified 102 landslide samples as positive (true positives) and 106 non-landslide samples as negative (true negatives). In addition, 43 non-landslide samples were incorrectly classified as landslides (false positives), while 47 landslide samples were incorrectly classified as non-landslides (false negatives).

These results indicate that the model produced misclassifications in both classes. False positives represent locations predicted to be susceptible to landslides that were labeled as non-landslide locations in the reference data. Conversely, false negatives represent landslide locations that the model failed to identify. The distribution of these errors provides a basis for assessing the model's ability to distinguish between the spatial characteristics of landslide and non-landslide locations.

ROC Curve

The receiver operating characteristic (ROC) curve illustrates the relationship between the true positive rate (TPR) and the false positive rate (FPR) across different classification thresholds applied to the model's predicted scores.

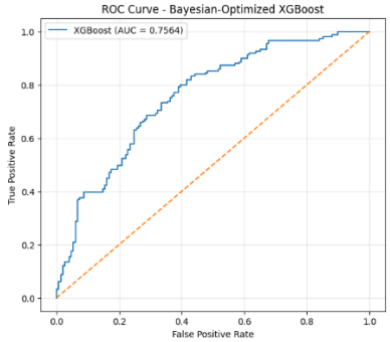


Figure 7. ROC Curve Model XGBoost

The evaluation results yielded an area under the curve (AUC) of 0.7564 for the XGBoost model. This value indicates the model's ability to discriminate between landslide and non-landslide classes across different classification thresholds.

The ROC curve lying above the reference line, which represents random classification, indicates that the model performs better than a random classifier. The AUC value suggests that the model learned predictive patterns associated with landslide conditioning factors in the study area.

SHAP Analysis Results

After the final XGBoost model was obtained, SHAP (SHapley Additive exPlanations) analysis was performed to examine the contributions of the eleven conditioning factors to the model's predictions.

SHAP analysis provides information about the magnitude and direction of each conditioning factor's contribution to the model output. The mean absolute SHAP value was used to assess the relative importance of each variable based on the average magnitude of its contribution to the model predictions, irrespective of direction.

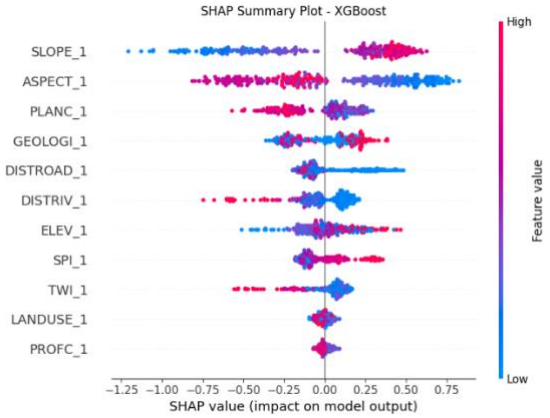


Figure 8. SHAP Summary Plot

The SHAP summary plot identified slope as the most influential conditioning factor, followed by aspect, plan curvature, geology, distance to road, distance to river, elevation, SPI, TWI, land use, and profile curvature. Slope contributed most to distinguishing landslide from non-landslide locations, while other topographic and hydrological factors also influenced model predictions. Positive SHAP values indicate contributions toward the landslide class, whereas negative values indicate contributions toward the non-landslide class. Feature colors represent their values, with red indicating higher and blue indicating lower values. Relative feature importance was assessed using mean absolute SHAP values, where higher values indicate greater contributions to model predictions.

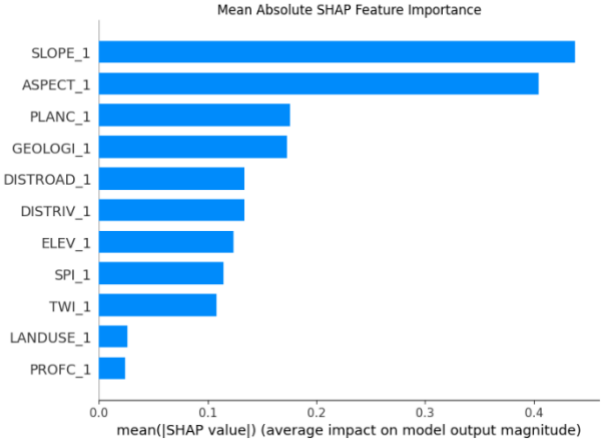


Figure 9. SHAP Feature Importance

The SHAP feature importance analysis identified slope as the most influential conditioning factor, with a mean absolute SHAP value of 0.438078, followed by aspect at 0.404144. Plan curvature and geology also made substantial contributions, with values of 0.175805 and 0.172908, respectively.

Other factors, including distance to road (0.133836), distance to river (0.133826), and elevation (0.123870), also contributed to the model's predictions. The hydrological indices SPI (0.114571) and TWI (0.107847) had comparatively lower contributions than the leading factors. Meanwhile, land use (0.026336) and profile curvature (0.023797) had the smallest mean absolute SHAP values.

Table 7
Contribution Ranking

Rank	Conditioning Factors	Mean SHAP
1	Slope	0,438078
2	Aspect	0,404144

3	Plan Curvature	0,175805
4	Geology	0,172908
5	Distance to Road	0,133836
6	Distance to River	0,133826
7	Elevation	0,123870
8	SPI	0,114571
9	TWI	0,107847
10	Land Use	0,026336
11	Profile Curvature	0,023797

The analysis indicates that the XGBoost model relied most heavily on information associated with topographic characteristics, particularly slope and aspect, when distinguishing between landslide and non-landslide locations. The substantial contributions of these factors suggest that variations in terrain steepness and orientation were important to the classification patterns learned by the model.

Direction of Influence Factors

The SHAP summary plot was used to examine the relationships between conditioning factor values and their contributions to the XGBoost model's predictions. A positive SHAP value indicates that a feature contributes to an increase in the model output for the landslide class, whereas a negative SHAP value indicates a contribution in the opposite direction.

Based on the mean absolute SHAP values, the five conditioning factors with the largest contributions were selected for further examination using SHAP dependence plots. These factors were slope, aspect, plan curvature, geology, and distance to road.

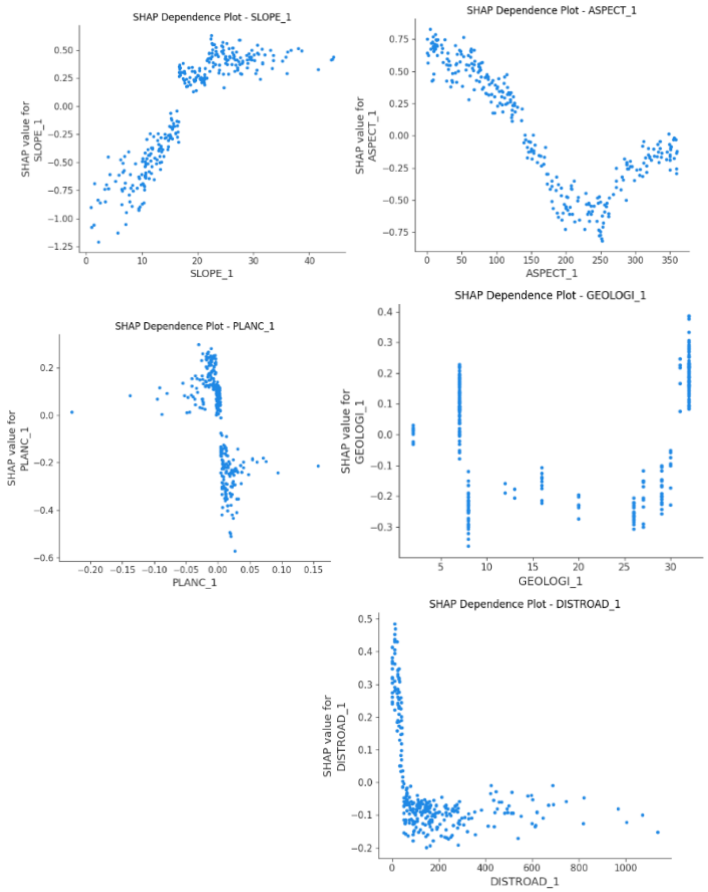


Figure 10. SHAP Dependence Plot Main Factors

The SHAP dependence plots were used to examine the relationships between individual conditioning factor values and their corresponding SHAP values. The results indicated that each factor exhibited a different relationship with the model's predicted landslide susceptibility.

For slope, SHAP values were negative at lower slope angles, approximately 0°–15°, and became positive at angles above approximately 20°. At higher slope angles, the SHAP values generally remained between approximately +0.3 and +0.5. This threshold-like pattern suggests that steeper slopes contributed to higher predicted landslide susceptibility within the fitted model.

SHAP analysis indicated higher landslide susceptibility on north- to east-facing slopes (0°–120°), concave terrain, and areas closer to roads. Conversely, south- to west-facing slopes (150°–260°) and flatter or convex terrain showed negative SHAP contributions. Geological classes exhibited varying contributions depending on lithology.

The directional patterns identified in the SHAP dependence plots are consistent with the dominance of topographic factors reported in previous studies. In Pacitan Regency, Mulabbi et al. (2025) identified slope, elevation, aspect, and distance to stream as the most influential factors, broadly agreeing with the importance of slope and aspect in the present model. This comparison supports the interpretation that topographic conditions contribute substantially to the spatial pattern of predicted susceptibility in the regency. However, the directional effects of individual factors should be interpreted in the context of local landforms, consistent with Sun et al. (2023), who emphasized the importance of landform differentiation.

Landslide Susceptibility Map

The XGBoost model with the optimal hyperparameter configuration was applied across the entire study area using raster layers representing the eleven landslide conditioning factors as spatial inputs. Predictions were generated for each pixel to obtain a model output representing the predicted likelihood of the landslide class.

Each pixel was assigned a predicted score between 0 and 1, representing the model's estimated likelihood of belonging to the landslide class. Scores closer to 0 indicated a lower predicted likelihood of landslide occurrence, whereas scores closer to 1 indicated a higher predicted likelihood.

The predicted scores were subsequently represented as a landslide susceptibility raster and visualized using QGIS to produce a landslide susceptibility map of Pacitan Regency.

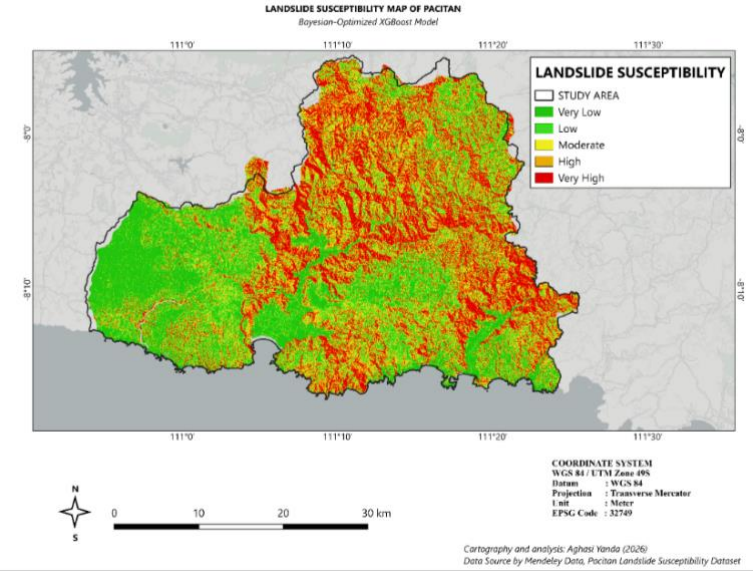


Figure 11. Map of Landslide Susceptibility of Pacitan Regency

The susceptibility raster was classified into five categories using the quantile classification method: Very Low, Low, Moderate, High, and Very High. This classification facilitated the interpretation of the spatial distribution of relative landslide susceptibility across the study area.

The resulting map revealed spatial variations in predicted landslide susceptibility across Pacitan Regency. These variations reflect differences in the combinations of conditioning factor values represented in the model, including topographic, geological, hydrological, land-use, and other environmental characteristics.

Key Results Summary

To provide a comprehensive overview of the study findings, the main results of landslide susceptibility modeling using Bayesian-optimized XGBoost and SHAP analysis are summarized in Table 8.

Table 8
Summary of the main results of the study

Components	Results
Dataset	1,486 sample points
Landslide	743 points
Non-landslide	743 points
Conditioning factors	11 factors
Dataset splitting	Training 80% (1,188 points), Testing 20% (298 points)
Hyperparameter optimization	Bayesian Optimization with TPE approach
Classification model	Extreme Gradient Boosting (XGBoost)
Best Mean CV ROC-AUC	0,7541
Best hyperparameters	n_estimators = 850; max_depth = 7; learning_rate = 0.0307; min_child_weight = 3; subsample = 0.7926; colsample_bytree = 0.8357; gamma = 3.9590; reg_alpha = 0.00685; reg_lambda = 8.16×10^{-6}
Accuracy	0,6980
Precision	0,7034
Recall	0,6846
F1 Score	0,6939
ROC-AUC	0,7564
The SHAP factor with the largest contribution	Slope (0.438078), Aspect (0.404144), Plan Curvature (0.175805), Geology (0.172908), Distance to Road (0.133836)
Map susceptibility	Distribution map of landslide susceptibility level in Pacitan Regency based on five classes (Very Low, Low, Moderate, High, Very High)

The results indicate that the XGBoost model optimized through Bayesian optimization achieved measurable performance in classifying landslide and non-landslide locations according to several evaluation metrics. SHAP analysis provided insights into the relative contributions of the conditioning factors used by the model, while the mapping process produced a spatial representation of relative landslide susceptibility across Pacitan Regency.

Discussion

Performance of the Bayesian-Optimized XGBoost Model

Bayesian optimization using the Tree-structured Parzen Estimator (TPE) identified an XGBoost hyperparameter configuration for landslide susceptibility modeling in Pacitan Regency. Using the mean cross-validation ROC-AUC as the optimization objective, the model achieved a value of 0.7541. The final model achieved an accuracy of 0.6980, precision of 0.7034, recall of 0.6846, F1-score of 0.6939, and ROC-AUC of 0.7564 on the testing dataset.

The ROC-AUC result indicates the model's ability to discriminate between landslide and non-landslide locations across classification thresholds. However, the confusion matrix revealed 47 false negatives and 43 false positives, indicating that classification errors remained. These errors may reflect the complexity of the relationships between environmental conditions and

landslide occurrence, as well as the limitations of relying on static conditioning factors. Bayesian optimization systematically selected hyperparameters based on previous evaluations (Snoek et al., 2012). Nevertheless, its superiority or computational efficiency relative to default XGBoost settings or alternative optimization methods cannot be established without direct comparative experiments.

Interpretation of Conditioning Factors Using SHAP

SHAP analysis (SHapley Additive exPlanations) offers an advantage over conventional feature importance measures by quantifying the magnitude and direction of each feature's contribution to model predictions at both the global and individual levels. The results indicate that slope had the largest mean absolute SHAP value (0.438078), followed by aspect (0.404144), plan curvature (0.175805), geology (0.172908), and distance to road (0.133836).

The dominance of slope and aspect in the SHAP analysis indicates that topographic characteristics contributed substantially to the classification patterns learned by the XGBoost model. From a geomorphological perspective, slope steepness is associated with the gravitational forces acting on soil and rock masses, whereas aspect is related to variations in environmental conditions, including solar exposure and moisture distribution. Plan curvature, geology, and distance to roads may also reflect relevant terrain, material, and anthropogenic characteristics. Nevertheless, SHAP values represent the contributions of variables to model predictions and do not independently establish the physical mechanisms responsible for landslide occurrence. Accordingly, these findings should be interpreted in conjunction with the local geomorphological context.

The SHAP dependence plots indicate that the relationships between conditioning factors and predicted susceptibility are not necessarily linear. This finding illustrates the ability of XGBoost to capture nonlinear relationships in the predictor data. The integration of XGBoost and SHAP therefore provides both spatial predictions of landslide susceptibility and interpretable information about the relationships between conditioning factors and the resulting classification outputs.

Spatial Interpretation of the Susceptibility Map

The resulting landslide susceptibility map presents the spatial distribution of predicted susceptibility across Pacitan Regency based on eleven environmental conditioning factors. The five susceptibility classes illustrate spatial variations, with higher scores indicating stronger associations between predictor characteristics and landslide occurrence in the trained model. By applying the model to all pixels across the study area, the map provides broader spatial information beyond the available landslide inventory points and supports the preliminary identification of susceptible areas for regional planning and disaster mitigation. However, since the model relies on static conditioning factors and a balanced classification dataset, the results represent relative spatial susceptibility rather than time-specific landslide probabilities or a comprehensive disaster risk assessment. Therefore, the susceptibility scores should be interpreted as indicators of spatial patterns, not predictions of when landslides will occur.

Research Implications and Limitations

Integrating Bayesian optimization, XGBoost, and SHAP enhances landslide susceptibility mapping through systematic hyperparameter tuning, nonlinear prediction, and model interpretability. The resulting susceptibility maps can support preliminary mitigation and regional planning. However, limitations include reliance on secondary data, static predictors, and uncertainty in non-landslide sample labeling(Samodra, 2024). The exclusion of dynamic factors, such as rainfall and soil moisture, and the non-causal nature of SHAP interpretations should also be considered. Future studies should incorporate dynamic variables and independent validation datasets to improve model reliability and generalizability.

CONCLUSION

This study developed an interpretable landslide susceptibility mapping framework for Pacitan Regency by integrating Bayesian optimization, XGBoost, and SHAP analysis. Using 1,486

samples comprising 743 landslide and 743 non-landslide locations, together with eleven environmental conditioning factors, the optimized XGBoost model achieved an accuracy of 69.80%, precision of 70.34%, recall of 68.46%, F1-score of 69.39%, and ROC-AUC of 0.7564 on the testing dataset. SHAP analysis identified slope, aspect, plan curvature, geology, and distance to road as the most influential predictors, highlighting the importance of topographic and environmental characteristics in the model's classification results.

The optimized model generated a landslide susceptibility map comprising five relative susceptibility classes, providing spatial information that may support preliminary disaster mitigation assessments and regional planning. The integration of Bayesian optimization, XGBoost, and SHAP combines systematic hyperparameter selection, predictive modeling, and interpretable feature contributions. However, classification errors, reliance on static environmental predictors, and uncertainty in non-landslide sample labels limit the model's generalizability. Future research should incorporate dynamic environmental variables and independent validation datasets to improve predictive reliability and support more comprehensive landslide susceptibility assessments.

ACKNOWLEDGEMENT

The authors acknowledge the availability of the landslide inventory and spatial datasets provided through Mendeley Data, which supported the landslide susceptibility modeling conducted in Pacitan Regency, Indonesia. The authors also recognize the contributions of the researchers whose published studies provided the theoretical and methodological foundations for this research.

AUTHOR CONTRIBUTION STATEMENT

Aghasi Yanda Wafa Azizah, Khalila Dzakia Agram, Shellomitha Sulvana Dewi, Saira Nashwa Ramadhan, and Yuliani Puji Astuti contributed to the research presented in this manuscript. The research activities included conceptualization, data preparation, spatial data processing, XGBoost modeling, Bayesian hyperparameter optimization, SHAP analysis, interpretation of findings, and manuscript preparation. The authors collectively contributed to the development and presentation of the research findings.

REFERENCES

Ado, M., Amitab, K., Maji, A. K., Jasińska, E., Gono, R., Leonowicz, Z., & Jasiński, M. (2022). Landslide susceptibility mapping using machine learning: A literature survey. *Remote Sensing*, *14*(13), 3029. https://doi.org/10.3390/rs14133029

Akiba, T., Sano, S., Yanase, T., Ohta, T., & Koyama, M. (2019). Optuna: A next-generation hyperparameter optimization framework. Dalam *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (2623–2631). https://doi.org/10.1145/3292500.3330701

Bergstra, J., Bardenet, R., Bengio, Y., & Kégl, B. (2011). Algorithms for hyper-parameter optimization. Dalam *Advances in Neural Information Processing Systems* (Vol. 24, hlm. 2546–2554).

Bergstra, J., & Bengio, Y. (2012). Random search for hyper-parameter optimization. *Journal of Machine Learning Research*, *13*, 281–305.

Bischl, B., Binder, M., Lang, M., Pielok, T., Richter, J., Coors, S., Thomas, J., Ullmann, T., Becker, M., Boulesteix, A.-L., Deng, D., & Lindauer, M. (2021). Hyperparameter optimization: Foundations, algorithms, best practices, and open challenges. *WIREs Data Mining and Knowledge Discovery*, *11*(2), e1484. https://doi.org/10.1002/widm.1484

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. Dalam *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (785–794). https://doi.org/10.1145/2939672.2939785

Darminto, M. R., Widodo, A., Alfatinah, A., & Chu, H. J. (2021). High-resolution landslide susceptibility map generation using machine learning (case study in Pacitan, Indonesia). *International Journal on Advanced Science, Engineering and Information Technology*, *11*(1),

369–379. https://doi.org/10.18517/ijaseit.11.1.11679

Dasgupta, S., & Sen, J. (2024). A comparative study of hyperparameter tuning methods. *Data Science in Theory and Practice*, 258.

Dimyati, M., Ash Shidiq, I. P., & Ichsandya, D. B. (2022). GIS-based binary logistic regression for landslide susceptibility mapping in the central part of Pacitan Regency, East Java, Indonesia. *International Journal on Advanced Science, Engineering and Information Technology*, *12*(4), 1462–1468. https://doi.org/10.18517/ijaseit.12.4.16221

Grandini, M., Bagli, E., & Visani, G. (2020). Metrics for multi-class classification: an overview. *arXiv preprint arXiv:2008.05756*.

Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., & Pedreschi, D. (2018). A survey of methods for explaining black box models. *ACM Computing Surveys*, *51*(5), Article 93. https://doi.org/10.1145/3236009

Hussain, M. A., Chen, Z., Shoaib, M., et al. (2022). Landslide susceptibility mapping using machine learning algorithm validated by persistent scatterer InSAR technique. *Sensors*, *22*(9), 3119. https://doi.org/10.3390/s22093119

Khan Inan, M. S., & Rahman, I. (2022). *Explainable AI integrated feature selection for landslide susceptibility mapping using TreeSHAP*. arXiv. https://doi.org/10.1007/s42979-023-01960-5

Kotsiantis, S. B., Kanellopoulos, D., & Pintelas, P. E. (2006). Data preprocessing for supervised learning. *International Journal of Computer Science*, *1*(2), 111–117.

Krawczyk, B. (2016). Learning from imbalanced data: Open challenges and future directions. *Progress in Artificial Intelligence*, *5*(4), 221–232. https://doi.org/10.1007/s13748-016-0094-0

Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., & Lee, S.-I. (2020). From local explanations to global understanding with explainable AI for trees. *Nature Machine Intelligence*, *2*(1), 56–67. https://doi.org/10.1038/s42256-019-0138-9

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. Dalam *Advances in Neural Information Processing Systems* (Vol. 30).

Molnar, C. (2020). *Interpretable machine learning*. Lulu. com.

Mulabbi, A., Danoedoro, P., & Samodra, G. (2025). Integrating interpretable artificial neural networks, geomorphic plausibility and transfer learning for landslide susceptibility mapping: A case study of Pacitan, East Java. *Discover Sustainability*, *6*(1). https://doi.org/10.1007/s43621-025-01959-3

Ponce-Bobadilla, A. V., Schmitt, V., Maier, C. S., Mensing, S., & Stodtmann, S. (2024). Practical guide to SHAP analysis: Explaining supervised machine learning model predictions in drug development. *Clinical and Translational Science*. https://doi.org/10.1111/cts.70056

Putra, A. N., Nita, I., Jauhary, M. R. A., Nurhutami, S. R., & Ismail, M. H. (2021). Landslide risk analysis on agriculture area in Pacitan Regency in East Java Indonesia using geospatial techniques. *Environment and Natural Resources Journal*, *19*(3). https://doi.org/10.32526/ennrj/19/2020167

Samodra, G., Malawani, M. N., Suhendro, I., & Mardiatno, D. (2024). Spatial datasets for benchmarking machine learning-based landslide susceptibility models. *Data in Brief*, *57*, 111155. https://doi.org/10.17632/vrtx3w6mjd.1

Snoek, J., Larochelle, H., & Adams, R. (2012). Practical bayesian optimization of machine learning algorithms. *Advances in neural information processing systems*, *25*.

Sun, D., Chen, D., Zhang, J., Mi, C., Gu, Q., & Wen, H. (2023). Landslide susceptibility mapping based on interpretable machine learning from the perspective of geomorphological differentiation. *Land*, *12*(5), Article 1018. https://doi.org/10.3390/land12051018

Watanabe, S. (2023). *Tree-structured Parzen estimator: Understanding its algorithm components and their roles for better empirical performance*. arXiv. https://arxiv.org/abs/2304.11127