

Comparative Analysis of Random Forest, SVM-RBF, and XGBoost with SHAP Interpretation for Google Earth Engine-Based Landslide Susceptibility Classification in Wonosobo Regency

Mizana Amaliya^{*1}, Erna Dwi Astuti², Rina Mahmudati³

^{1,2,3} Department of Informatics Engineering, Quranic Science University, Wonosobo, Indonesia
Email: ¹mizanaamaliya@mhs.unsiq.ac.id, ²ernada@unsiq.ac.id, ³rinamahmud@unsiq.ac.id

Abstrak

Tanah longsor merupakan bencana dominan di Kabupaten Wonosobo, Jawa Tengah, dengan 190 kejadian (62% dari total bencana) pada tahun 2024. Hasil penelitian terdahulu menunjukkan kinerja algoritma *machine learning* yang tidak konsisten akibat perbedaan wilayah, data, fitur, dan skema evaluasi. Berdasarkan penelusuran literatur, belum ditemukan evaluasi komparatif Random Forest (RF), Support Vector Machine (SVM-RBF), dan XGBoost yang memadukan *Google Earth Engine* (GEE), interpretasi SHAP, dan klasifikasi tiga kelas kerawanan untuk Kabupaten Wonosobo. Penelitian ini bertujuan membandingkan kinerja ketiga algoritma dalam mengklasifikasikan kerawanan longsor berbasis data geospasial. Sebanyak 642 sampel (321 titik longsor dan 321 non-longsor) digunakan dengan delapan faktor kondisional yang diekstraksi melalui GEE, dengan tujuh faktor dipertahankan setelah pengujian VIF. Model dievaluasi menggunakan akurasi, *Cohen's Kappa*, *Matthews Correlation Coefficient* (MCC), dan AUC-ROC dengan *10-fold stratified cross-validation*. XGBoost menunjukkan kinerja tertinggi pada dataset penelitian dengan akurasi 93,02%, AUC-ROC 0,974, dan *cross-validation accuracy* $91,0\% \pm 3,1\%$. Analisis SHAP menunjukkan bahwa curah hujan memberikan kontribusi terbesar terhadap prediksi model, yang menunjukkan pentingnya fitur tersebut dalam keputusan model. Klasifikasi probabilitas prediksi pada titik sampel menggunakan *Jenks Natural Breaks* menghasilkan 49,2% kelas Rendah, 6,4% Sedang, dan 44,4% Tinggi.

Kata kunci: geospasial, kerawanan longsor, *machine learning*, wonosobo, XGBoost

Abstract

Landslides constitute the predominant disaster type in Wonosobo Regency, Central Java, with 190 recorded events (62% of all disaster incidents) in 2024. Prior studies have reported inconsistent performance among machine learning algorithms due to variations in geographic context, dataset characteristics, feature selection, and evaluation frameworks. A review of the literature reveals a lack of comparative evaluations of Random Forest (RF), Support Vector Machine with Radial Basis Function kernel (SVM-RBF), and XGBoost that integrate Google Earth Engine (GEE), SHAP interpretability analysis, and three-class susceptibility classification specifically for Wonosobo Regency. This study aims to evaluate and compare the performance of these three algorithms in mapping landslide susceptibility using geospatial data. A total of 642 sample points (321 landslide and 321 non-landslide locations) were analyzed using eight conditioning factors extracted via GEE, seven of which were retained following Variance Inflation Factor (VIF) testing. Model performance was evaluated using accuracy, Cohen's Kappa, Matthews Correlation Coefficient (MCC), and AUC-ROC with 10-fold stratified cross-validation. XGBoost achieved the highest performance on the study dataset, with an accuracy of 93.02%, an AUC-ROC of 0.974, and a cross-validation accuracy of $91.0\% \pm 3.1\%$. SHAP analysis indicated that rainfall made the largest contribution to the model predictions, highlighting its importance in the model's decision-making process. Classification of prediction probabilities at the sample points using the Jenks Natural Breaks method yielded a susceptibility distribution of 49.2% Low, 6.4% Moderate, and 44.4% High.

Keywords: geospatial, landslide susceptibility, machine learning, Wonosobo, XGBoost

This work is an open access article and licensed under a Creative Commons Attribution-NonCommercial ShareAlike 4.0 International (CC BY-NC-SA 4.0)



1. INTRODUCTION

Landslides represent a recurring geological hazard in tropical and mountainous regions, causing substantial losses ranging from fatalities to infrastructure damage and agricultural land degradation [1]. As a country characterized by mountainous topography, high rainfall intensity, and active tectonic

activity, Indonesia is particularly susceptible to this threat [2]. According to data from the National Disaster Management Authority (Badan Nasional Penanggulangan Bencana/BNPB), hydrometeorological disasters including landslides account for more than 90% of all disaster occurrences in Indonesia. In 2024 alone, a total of 5,593 national disaster events were recorded, of which 933 were landslide incidents distributed across various regions [3]. This trend has persisted into subsequent years, with landslides reported as the deadliest natural disaster affecting communities throughout 2025 [4].

Wonosobo Regency, Central Java, is among the regions with the highest landslide susceptibility in Indonesia. This vulnerability is primarily driven by its topographic characteristics, which consist of hilly and mountainous terrain ranging in elevation from 250 to 2,250 meters above sea level [5]. The combination of geological, climatological, hydrological, and geographical conditions renders Wonosobo Regency particularly prone to landslide occurrences [6]. Compounding this risk are land-use changes driven by increasing land demand and persistently high rainfall, both of which significantly elevate landslide potential [7]. Data from the Regional Disaster Management Agency (Badan Penanggulangan Bencana Daerah/BPBD) of Wonosobo Regency indicate that 303 disaster events were recorded in 2024, of which 190 incidents approximately 62% were attributed to landslides [8], with total economic losses from all natural disasters reaching IDR 2,623,300,000 (BPBD Wonosobo Regency, personal communication, 2026). These figures underscore the critical need for accurate and adaptive spatial data-based landslide susceptibility mapping to effectively support disaster mitigation decision-making.

Landslide susceptibility mapping and classification constitute a critical component of disaster risk reduction, as they provide spatially accurate information on hazard levels across a given area [9]. Through the integration of Geographic Information Systems (GIS) with remote sensing data, key landslide-triggering parameters including slope gradient, rainfall, land use, and soil type can be comprehensively analyzed within a location-based framework [10]. However, the relationships among these variables are often nonlinear and difficult to model using conventional approaches. Consequently, machine learning has emerged as an increasingly relevant alternative, offering the capacity to process multivariate data, detect latent patterns, and generate more accurate and adaptive predictions in response to environmental dynamics [11]. This approach provides the analytical foundation for landslide susceptibility classification based on spatial data in the present study.

Various machine learning algorithms have been applied to landslide susceptibility mapping, but their performance varies across study areas and datasets. In Inje, South Korea, using 10 conditioning factors and 28,754 balanced samples, XGBoost slightly outperformed RF and SVM, with accuracies of 0.947, 0.941, and 0.918, respectively [12]. In Darjeeling Kalimpong, India, using 21 factors and 1,482 samples, RF, XGBoost, and SVM achieved similar accuracies, while XGBoost obtained the highest AUC of 0.987 [13]. Conversely, in the Upper Jinsha River, China, using 18 factors and 163 inventory points, RF outperformed XGBoost and SVM with an accuracy of 0.793 [14]. RF also outperformed SVM in Ningqiang, China, with accuracies of 78% and 71%, respectively [15]. However, the comparative performance of RF, SVM-RBF, and XGBoost using the same geospatial dataset and evaluation framework has not been sufficiently investigated in Wonosobo Regency. This comparison is important because Wonosobo is characterized by steep and mountainous terrain, high rainfall, and ongoing land-use changes, resulting in spatially heterogeneous landslide susceptibility and potentially complex nonlinear relationships among conditioning factors that may be difficult to capture using conventional approaches.

The selection of RF, SVM-RBF, and XGBoost in this study was based on how well each algorithm's characteristics matched the nature of the landslide geospatial data in Wonosobo Regency, which features non-normal distribution, nonlinear relationships among triggering variables, and potential multicollinearity among spatial factors. Random Forest (RF), Support Vector Machine (SVM) with a Radial Basis Function (RBF) kernel, and XGBoost were chosen because each offers a distinct algorithmic approach to handling geospatial classification problems. RF is well-regarded for its ability to handle noisy data through a bagging-based ensemble approach that combines multiple decision trees to improve model accuracy and stability [16]. SVM-RBF, on the other hand, excels at recognizing complex nonlinear patterns by mapping data into a high-dimensional feature space via the kernel trick, making the data more linearly separable without suffering from the curse of dimensionality [17]. XGBoost, meanwhile, delivers strong predictive performance through a gradient boosting mechanism

equipped with regularization, which enhances model generalization and reduces the risk of overfitting [18]. Prior to modeling, all geospatial data underwent a standardized preprocessing pipeline to ensure that the data quality fed into each algorithm remained consistent, enabling a fair and comparable evaluation. With this approach, any differences in model performance can be attributed to the inherent characteristics of each algorithm, making the comparison more objective and equitable.

Therefore, this study aims to evaluate and compare the performance of RF, SVM-RBF, and XGBoost for geospatial-based landslide susceptibility classification in Wonosobo Regency using accuracy, Cohen's Kappa, MCC, and AUC-ROC. The study also aims to identify the relative contribution of conditioning factors through SHAP analysis and classify landslide susceptibility at sampled locations based on the predicted probabilities of the best-performing model.

The contributions of this study are threefold: (1) providing a comparative evaluation of RF, SVM-RBF, and XGBoost under the same geospatial dataset and evaluation framework; (2) improving model interpretability by quantifying the magnitude and direction of conditioning factors using SHAP analysis; and (3) classifying landslide susceptibility into three classes at sampled locations based on the predicted probabilities of the best-performing model.

2. RESEARCH METHODOLOGY

2.1 Research Framework

This study consists of six sequential stages: data collection, data preprocessing, classification modeling, model evaluation, feature interpretation, and susceptibility classification, as illustrated in Figure 1. The output of each stage is used as the input for the subsequent stage to obtain the final landslide susceptibility classification.

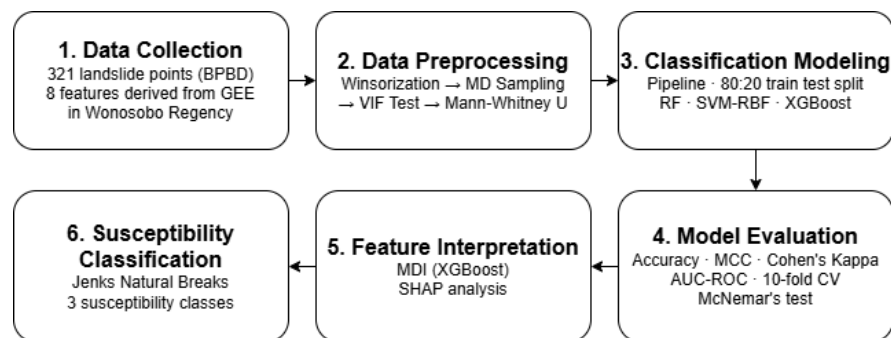


Figure 1. Research Framework for Landslide Susceptibility Classification in Wonosobo Regency

2.2 Study Area and Data

The study area encompasses Wonosobo Regency, Central Java, situated at elevations ranging from 250 to over 2,250 meters above sea level, characterized by a predominantly hilly terrain and steep slopes [19]. Landslide occurrence data were obtained from the Wonosobo Regency Regional Disaster Management Agency (BPBD), comprising verified landslide coordinate points recorded during the 2015-2025 period. These points serve as presence data (landslide class) within the binary dataset constructed for this study.

Geospatial data used as conditional variables were extracted via the Google Earth Engine (GEE) platform at a spatial resolution of 30 meters. The eight conditional features employed in this study and their respective data sources are presented in Table 1. The selection of these features was based on a review of the literature concerning conditional factors associated with landslide occurrence [20]. GEE is a cloud-based geospatial computing platform developed by Google, designed to efficiently process and analyze large-scale remote sensing data without the need for intensive local infrastructure [21]. GEE provides access to a global catalog of satellite and geospatial data spanning more than 40 years of historical records, including imagery from Landsat, Sentinel, MODIS, SRTM, and various climate datasets such as CHIRPS [22].

Table 1. Geospatial Features Used in This Study

No	Feature	Data Source (GEE)	Description
1	Elevation (m)	SRTM DEM 30 m	Terrain elevation above sea level

2	Slope (degrees)	SRTM DEM derivative	Slope gradient influencing the rate of mass movement
3	Curvature	SRTM DEM derivative	Slope curvature affecting flow convergence or divergence
4	TWI	SRTM DEM 30 m	Topographic Wetness Index: soil moisture accumulation index
5	Rainfall (mm/day)	CHIRPS	Mean daily rainfall for the period 2015-2025
6	Dist_Roads (m)	OSM via GEE	Euclidean distance to the nearest road network
7	Dist_Rivers (m)	OSM via GEE	Euclidean distance to the nearest river network
8	Landuse	ESA WorldCover 10 m	Land cover class influencing soil water absorption and surface erosion

Within this study, GEE served as the primary platform for extracting all geospatial conditional features used in landslide susceptibility modeling across Wonosobo Regency. The data extraction process was carried out through several sequential steps within the GEE environment. First, the administrative boundary of Wonosobo Regency was imported as the Region of Interest (ROI) defining the spatial extent of the analysis. Second, each satellite dataset was accessed using the Image Collection function in GEE, a cloud-based platform that provides free access to petabyte-scale spatial data for global environmental monitoring [23]. All datasets were subsequently processed at a uniform spatial resolution of 30 meters using the reproject function to the WGS84 coordinate reference system (EPSG:4326), ensuring harmonization across layers derived from multiple sensor sources [24]. The SRTM 30 m DEM was used to derive morphometric features, including slope and curvature, via the ee.Terrain function. Daily rainfall data were obtained from the Climate Hazards Group InfraRed Precipitation with Station data (CHIRPS) collection and averaged over the selected study period. Road and river networks were obtained from OpenStreetMap (OSM) via GEE, and Euclidean distances to these vector features were computed using the distance() function to determine proximity radii. Finally, land cover data were sourced from the ESA WorldCover 10 m product, reprojected and resampled to a uniform 30 m resolution to maintain pixel dimension consistency throughout the spatial analysis [24].

Feature values at each sample point were extracted using the sampleRegions() function in GEE, which enables simultaneous pixel value retrieval at both landslide and non-landslide coordinate points across all feature layers in a single operation. The extracted data were subsequently exported in comma-separated values (CSV) format via the Export.table.toDrive() function to Google Drive, and then processed within a Python environment on Google Colaboratory. This GEE-based approach ensures spatial resolution consistency, uniform temporal coverage, and reproducibility of the data acquisition process at the regency scale.

Table 2. Representative Records of the Geospatial Dataset Used in This Study

Curvature	Dist_Rivers (m)	Dist_Roads (m)	Elevation (m)	Landuse	Rainfall (mm/day)	Slope (°)	TWI	Class
20.7636	6.79	44.11	1583	10	8.75	5	9	1
6.5811	10.87	8.42	1373	10	10.68	7	8	1
9.3491	11.48	6.65	1367	40	10.68	6	9	1
13.1698	11.16	6.42	934	40	5.94	2	7	0
13.1701	25.26	6.65	780	10	5.94	2	7	0

Table 2 presents several representative records from the geospatial dataset used in this study. These samples illustrate the dataset structure and variables employed in the analysis.

2.3 Data Preprocessing

Data preprocessing is a fundamental stage in the model development cycle aimed at improving the quality of raw data to ensure readiness for modeling. Given that raw data are frequently susceptible

to noise, inconsistencies, and incompleteness, this stage encompasses data cleaning, transformation, and reduction or selection techniques to minimize errors and ensure greater data representativeness, thereby improving predictive accuracy [25]. Data preprocessing in this study was conducted through five sequential stages:

- Data Cleaning, comprising the removal of non-feature columns, standardization of numeric formats, and elimination of records containing missing values.
- Winsorization, a statistical preprocessing technique for handling outliers by replacing extreme values (excessively high or low) with defined boundary values, without discarding observations [26].
- Mahalanobis Distance (MD) filtering, applied to ensure that selected non-landslide points exhibit geospatial characteristics maximally contrasting with presence data. This approach aims to establish sharper class separation, thereby enhancing the discriminative capacity of the model [27].
- Multicollinearity Testing using the Variance Inflation Factor (VIF), whereby features with VIF values exceeding 10 were iteratively eliminated.
- Statistical Testing to assess data distribution. Normality was evaluated using the Shapiro-Wilk test ($\alpha = 0.05$). In cases where normality assumptions were violated, inter-class differences were examined using the Mann-Whitney U test to identify features with statistically significant differences between landslide and non-landslide classes [28].

2.4 Landslide Susceptibility Classification

This study classifies landslide susceptibility into three categories: Low, Moderate, and High. The classification approach is based on predicted landslide class probability values derived from the best-performing model identified through comparative evaluation. The probability values generated by the model reflect its confidence regarding the likelihood of a given location belonging to the landslide class, and are therefore used as a proxy for susceptibility levels.

Class boundaries were determined using the Jenks Natural Breaks method, applied to the distribution of predicted probability values across the entire dataset. The Jenks method was selected for its data-driven classification capability, operating by minimizing within-class variance and maximizing between-class variance. This results in more accurate class delineation by conforming to the natural distribution pattern of data within the study area [29]. Through this approach, landslide susceptibility levels can be mapped objectively in accordance with actual field conditions, following the classification scheme presented in Table 3.

Table 3. Three-Class Landslide Susceptibility Classification Scheme

Class	Susceptibility Level	Probability Score Range
3	High	$H > B_2$ (Jenks Break 2)
2	Moderate	$B_1 < H \leq B_2$ (Jenks Break 1 - Break 2)
1	Low	$H \leq B_1$ (Jenks Break 1)

2.5 Classification Modeling

This study compares three machine learning classification algorithms for landslide susceptibility mapping. Random Forest (RF) is a bagging-based ensemble method that constructs multiple decision trees and determines predictions through majority voting, effectively reducing overfitting while providing feature importance estimates [30]. Support Vector Machine (SVM) is a machine learning algorithm grounded in structural risk minimization, which seeks an optimal separating hyperplane that maximizes the margin between classes. For nonlinear relationships, SVM employs the kernel trick such as the Gaussian Radial Basis Function (RBF) kernel to project data into a higher-dimensional feature space [31]. XGBoost is a gradient boosting method that builds models iteratively to minimize a loss function, incorporating regularization to achieve high performance on structured data [32]. All models were implemented in Python 3.13.15 on Google Colaboratory using scikit-learn 1.6.1 (RandomForestClassifier, SVC, Pipeline, and GridSearchCV), XGBoost 3.4.1 (XGBClassifier), SHAP 0.52.0, statsmodels 0.15.0, jenksPy 0.4.1, NumPy 2.1.3, pandas 2.2.3, and SciPy 1.16.3. A fixed random seed of 42 was used for reproducibility.

2.6 Pipeline and Hyperparameter Tuning

All preprocessing and modeling stages were integrated into a unified workflow using the Pipeline class from the Scikit-Learn library to prevent data leakage [33]. The pipeline consists of two sequential components: (1) StandardScaler, which normalizes features to a distribution with zero mean and unit variance; and (2) the classification model. By employing an integrated pipeline, normalization parameters are calibrated exclusively on training data within each fold, thereby ensuring that no information from the test set propagates into the training process.

Data were partitioned using a stratified 80:20 split (513 training and 129 test samples). Hyperparameter optimization was performed on the training set using GridSearchCV with five-fold Stratified K-Fold cross-validation and weighted F1-score. The search spaces are presented in Table 4, comprising 12, 12, and 16 configurations for RF, SVM-RBF, and XGBoost, respectively. Unlisted parameters used library defaults, while XGBoost used fixed values of 1.0, 0, and 1 for `colsample_bytree`, `reg_alpha`, and `reg_lambda`, respectively. SVM used `probability = True` for AUC-ROC calculation.

Table 4. Hyperparameter search space

Hyperparameter	RF	SVM-RBF	XGBoost
n_estimators	{100, 200}	-	{100, 200}
max_depth	{10, 20, None}	-	{4, 6}
min_samples_leaf	{1, 2}	-	-
learning_rate	-	-	{0.05, 0.1}
subsample	-	-	{0.8, 1.0}
C	-	{0.01, 0.1, 0.5, 1}	-
gamma	-	{scale, auto, 0.01}	-
kernel	-	{rbf}	-

2.7 Model Evaluation

Model performance was assessed using accuracy, precision, recall, and F1-score (macro-averaged) to represent overall performance across all classes, as well as the Matthews Correlation Coefficient (MCC) to account for potential class imbalance [34][35]. Discriminative capacity was evaluated using the macro-averaged AUC-ROC. To ensure result reliability, a ten-fold stratified cross-validation scheme was applied to the complete dataset. Additionally, McNemar's test with Yates' continuity correction ($\alpha = 0.05$) was used to determine whether statistically significant performance differences existed among the compared algorithms. The evaluation metrics employed in this study are formally defined as follows.

Accuracy measures the proportion of correct predictions relative to all predictions and is formulated in Equation (1). Precision measures the proportion of correctly predicted positive instances, as expressed in Equation (2). Recall measures the model's ability to detect all actual positive cases, as given in Equation (3). The F1-score is the harmonic mean of precision and recall, as defined in Equation (4).

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \quad (1)$$

$$Precision = \frac{TP}{(TP + FP)} \quad (2)$$

$$Recall = \frac{TP}{(TP + FN)} \quad (3)$$

$$F1 - score = \frac{2 \times (Precision \times Recall)}{(Precision + Recall)} \quad (4)$$

Because the classification task consists of two classes, precision, recall, and F1-score were calculated for each class and then macro-averaged by taking the unweighted arithmetic mean across the two classes. The macro-averaged metrics are defined as follows:

$$P_{macro} = \frac{1}{K} \sum_{k=1}^K P_k \quad (5)$$

$$R_{macro} = \frac{1}{K} \sum_{k=1}^K R_k \quad (6)$$

$$F1_{macro} = \frac{1}{K} \sum_{k=1}^K F1_k \quad (7)$$

Where (K) denotes the total number of classes. Since this study involves binary classification, (K=2).

The Matthews Correlation Coefficient (MCC), presented in Equation (8), is a more robust metric for imbalanced data, as it accounts for all elements of the confusion matrix (TP, TN, FP, FN) [34]. MCC values range from -1 (perfect inverse prediction) to +1 (perfect prediction), with a value of 0 indicating random prediction.

$$MCC = \frac{(TP \times TN - FP \times FN)}{\sqrt{((TP+FP)(TP+FN)(TN+FP)(TN+FN))}} \quad (8)$$

Cohen's Kappa, defined in Equation (9), quantifies the level of agreement between model predictions and actual labels, accounting for the possibility of agreement occurring by chance. Kappa values approaching 1 indicate very high agreement between predictions and ground truth.

$$Kappa = \frac{(Po - Pe)}{(1 - Pe)} \quad (9)$$

where Po is the observed agreement proportion and Pe is the expected agreement proportion under chance.

2.8 Feature Importance and SHAP Analysis

The contribution of geospatial features to the best-performing model was assessed using two complementary approaches. First, gain-based feature importance from XGBoost was used to quantify the contribution of each feature based on the improvement in the model's objective function across the trees. Second, SHAP (SHapley Additive exPlanations) analysis using TreeExplainer was employed to quantify the contribution of each feature to individual predictions based on cooperative game theory. A beeswarm SHAP plot was used to visualize the distribution of SHAP values across the test dataset and the directional influence of feature values on landslide class predictions.

3. RESULTS AND DISCUSSION

3.1. Data Exploration and Preprocessing

Data preprocessing commenced with the cleaning of raw geospatial data and non-landslide sample selection based on Mahalanobis Distance (MD). From 321 verified landslide occurrence points, 321 non-landslide points exhibiting the most geospatially contrasting characteristics were selected, yielding a balanced dataset of 642 samples (321 landslide + 321 non-landslide). Outliers were subsequently handled using Winsorization by limiting values to the 5th and 95th percentiles.

Feature selection was subsequently performed to ensure data quality prior to modeling. The Variance Inflation Factor (VIF) analysis revealed that the Topographic Wetness Index (TWI) yielded a VIF value of 11.82 (VIF > 10) and was therefore removed to avoid multicollinearity and information redundancy. Following the removal of TWI, the remaining seven geospatial features curvature, dist_rivers, dist_roads, elevation, landuse, rainfall, and slope all exhibited VIF values below 10. This finding was further corroborated by Spearman correlation analysis, which indicated weak inter-feature associations, with the highest correlation coefficient of only 0.23 observed between elevation and landuse.

After multicollinearity was assessed, significance testing was conducted to evaluate differences in feature distributions between the landslide and non-landslide classes. Given the non-normal feature distributions identified by the Shapiro–Wilk test, the Mann-Whitney U test ($\alpha = 0.05$) was applied.

Results indicated that four features slope ($p < 0.0001$), dist_roads ($p = 0.0001$), rainfall ($p = 0.0025$), and curvature ($p = 0.0144$) exhibited statistically significant inter-class differences.

3.2. Comparative Evaluation of Classification Models

A comparative evaluation was conducted across three machine learning algorithms Random Forest (RF), Support Vector Machine with RBF kernel (SVM-RBF), and XGBoost using a ten-fold stratified cross-validation scheme. All models were optimized through GridSearchCV with an inner five-fold cross-validation. The comparative evaluation results are presented in Table 5.

Table 5. Comparative Performance of Landslide Susceptibility Classification Models

Model	Test Acc	Kappa	MCC	AUC-ROC	CV Acc ($\pm$ Std)
Random Forest	0.837	0.674	0.674	0.924	0.869 $\pm$ 0.051
SVM-RBF	0.760	0.520	0.528	0.834	0.723 $\pm$ 0.056
XGBoost*	0.930	0.861	0.861	0.974	0.910 $\pm$ 0.031

*Best-performing model; Acc = Accuracy; CV Acc = Cross-validation Accuracy

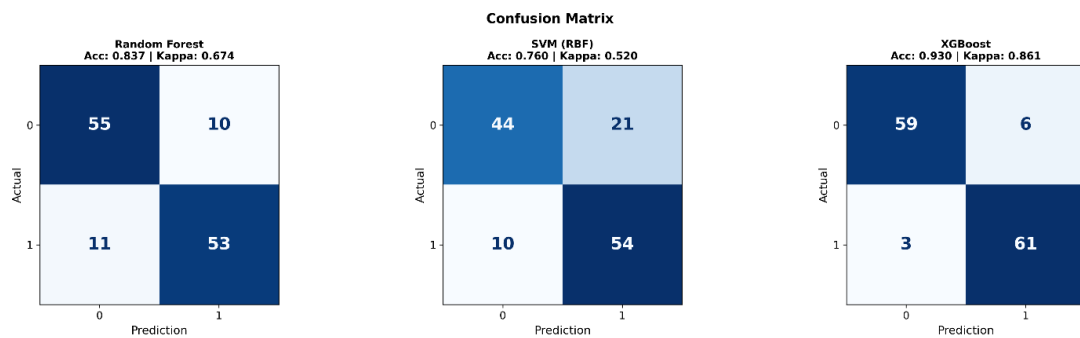


Figure 2. Confusion Matrix Comparison Across Models

As shown in Table 5, XGBoost obtained the highest metric values among the three models, with a test accuracy of 93.02%, Cohen's Kappa of 0.861, MCC of 0.861, and AUC-ROC of 0.974. The cross-validation accuracy of XGBoost was $91.0\% \pm 3.1\%$.

Random Forest ranked second with a test accuracy of 83.72% and an AUC-ROC of 0.924. Despite its competitive performance, Random Forest exhibited significant overfitting, with a perfect train accuracy of 1.000 and a gap of 0.163 between training and test accuracy. SVM-RBF recorded the lowest performance, with a test accuracy of 75.97% and an AUC-ROC of 0.834.

The confusion matrices presented in Figure 2 showed that XGBoost produced 6 false positives and 3 false negatives out of 129 test samples, while Random Forest produced 10 false positives and 11 false negatives. SVM-RBF recorded the highest number of false positives (21) and 10 false negatives.

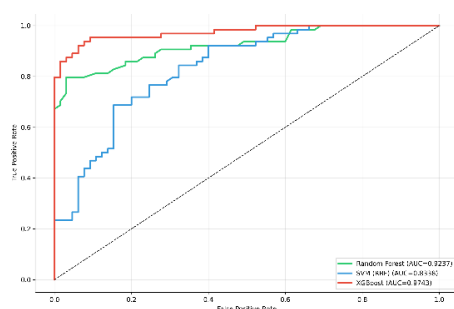


Figure 3. ROC Curves for Landslide Susceptibility Classification Model Comparison

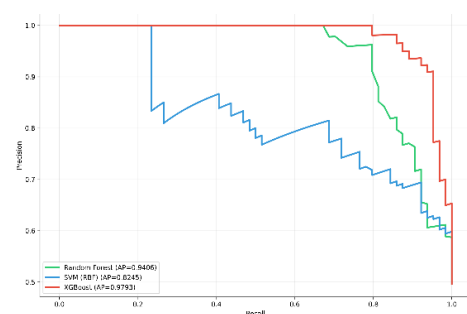


Figure 4. Precision-Recall Curves for Classification Model Comparison

Based on the evaluation results, XGBoost achieved the highest values among the evaluated models for accuracy, Kappa, MCC, and AUC-ROC, and was therefore selected as the primary model for subsequent feature interpretation analysis.

3.3. Feature Importance and SHAP Analysis

Feature contribution analysis was performed using two complementary approaches: Mean the gain-based feature importance based on XGBoost and SHAP (SHapley Additive exPlanations) analysis. Based on the gain-based feature importance scores presented in Figure 5, rainfall emerged as the most influential feature with an importance score of 0.221, followed by curvature (0.150), dist_roads (0.143), elevation (0.133), slope (0.127), dist_rivers (0.127), and landuse (0.100). The relatively uniform distribution of importance scores across features (ranging from 0.100 to 0.221) suggests that all features contribute meaningfully to the model's decision-making process, without excessive dominance by any single feature.

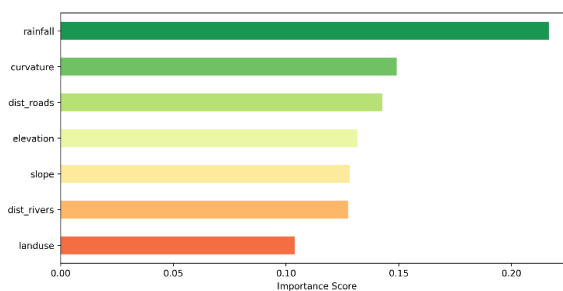


Figure 5. Gain-Based Feature Importance from XGBoost

SHAP analysis provided a theoretically more rigorous and interpretively richer perspective. The computed mean absolute SHAP values are presented in Table 6. Rainfall retained the top ranking with a mean absolute SHAP value of 1.658, substantially exceeding all other features, indicating that rainfall made the largest contribution to the model predictions. It was followed by dist_roads with a value of 0.690 and slope with a value of 0.627.

Table 6. SHAP Feature Importance Rankings for the XGBoost Model

Rank	Feature	Mean Absolute SHAP Value
1	rainfall	1.658
2	dist_roads	0.690
3	slope	0.627
4	elevation	0.511
5	curvature	0.406
6	dist_rivers	0.340
7	landuse	0.031

While mean absolute SHAP values indicate the magnitude of each feature's contribution, the direction of the contribution is interpreted from the distribution of SHAP values. The second and third rankings were occupied by dist_roads (0.690) and slope (0.627), respectively. The significant influence of road proximity is attributed to road construction on sloped terrain, which alters drainage patterns, causes slope undercutting, and introduces dynamic loading all of which elevate landslide risk. Slope, as a classical morphometric variable, has been consistently identified as a critical factor across numerous landslide susceptibility studies [9][14].

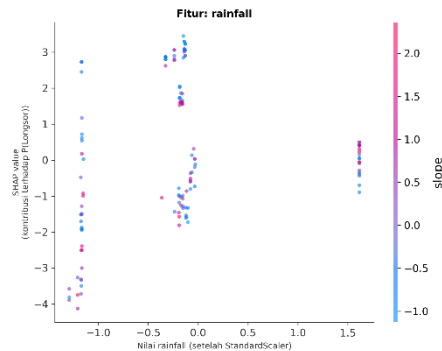


Figure 6. SHAP Dependence Plot for Rainfall (XGBoost Model)

The SHAP dependence plot (Figure 6) reveals a nonlinear interaction between rainfall and slope. At high rainfall values (above the mean post-StandardScaler normalization), SHAP values tend to be large and positive (approximately 1.5 to 3.5), particularly when slope values are also high (as indicated by the pink-to-red color scale). This pattern indicates that higher rainfall values contribute positively to the predicted landslide class, particularly under steeper slope conditions. Conversely, under low rainfall conditions, SHAP values are predominantly negative (down to -4.0), indicating a contribution toward a lower predicted landslide risk.

Beyond the interaction between rainfall and slope, the SHAP analysis also shows meaningful contributions from road proximity (dist_roads) and elevation. The relatively high mean SHAP value for dist_roads (0.690) indicates that locations closer to road networks tend to have a higher predicted probability of landslides. This finding is consistent with the theory that road construction in sloped areas can disrupt natural drainage patterns and introduce lateral stress on slopes, thereby reducing slope stability and increasing landslide risk [36]. Elevation (mean |SHAP| = 0.511) also provides a relatively substantial positive contribution, possibly because areas at higher elevations in the mountainous region of Wonosobo tend to be associated with steeper slopes. Interestingly, landuse recorded the lowest SHAP value (0.031).

3.4. Three-Class Landslide Susceptibility Classification

Landslide occurrence probabilities for all 642 dataset points were computed using the best-performing XGBoost model. These probability values were subsequently classified into three susceptibility levels using the Jenks Natural Breaks method, yielding threshold values of 0.326 for the first break and 0.835 for the second break. The distribution of predicted probabilities across each susceptibility class is illustrated in Figure 7, and a summary of the classification results is provided in Table 7.

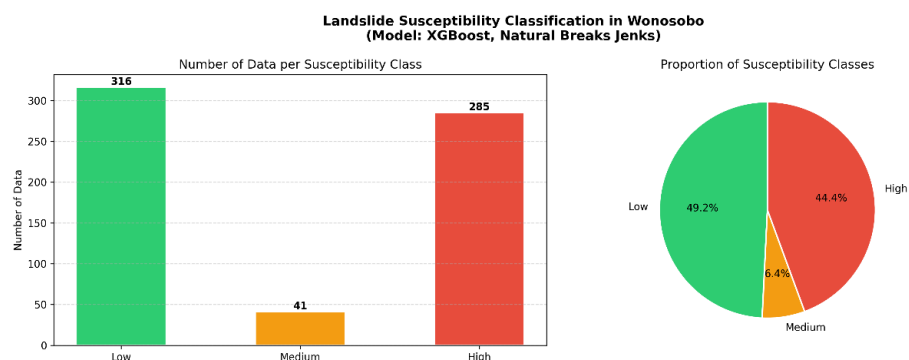


Figure 7. Landslide Probability Distribution by Susceptibility Class (XGBoost Model)

Table 7. Landslide Susceptibility Class Distribution Based on Jenks Natural Breaks

Susceptibility Class	Probability Range	Number of Points	Proportion (%)
Low	$P \leq 0.326$ (Jenks Break 1)	316	49.2%

Moderate	$0.326 < P \leq 0.835$ (Jenks Break 2)	41	6.4%
High	$P > 0.835$	285	44.4%

The classification results indicate that the majority of locations fall within the Low susceptibility class (316 points; 49.2%) and the High susceptibility class (285 points; 44.4%), while the Moderate class accounts for only 41 points (6.4%).

It should be noted that although this three-class classification framework is valuable for disaster mitigation planning purposes, the Jenks-derived thresholds are data-driven and depend on the specific distribution of the dataset used in this study. Application to new data or different geographical areas would require threshold recalibration to ensure the consistency and validity of the classification.

3.5. Statistical Significance Testing Across Models

McNemar's test with Yates' continuity correction ($\alpha = 0.05$) was conducted to assess the statistical significance of performance differences among the models based on their predictions on the test set. The results showed that XGBoost differed significantly from both Random Forest ($p < 0.05$) and SVM-RBF ($p < 0.01$). In contrast, the difference between Random Forest and SVM-RBF was not statistically significant ($p = 0.681$).

3.6. Discussion

XGBoost achieved the highest metric values under the dataset and evaluation scheme used in this study. This performance may be related to the 642 samples and seven geospatial features, which showed significant differences between landslide and non-landslide samples for slope, dist_roads, rainfall, and curvature. The combination of these geospatial predictors may involve nonlinear interactions that can be captured by tree-based models. The XGBoost configuration was optimized using GridSearchCV, including n_estimators, max_depth, learning_rate, and subsample, allowing the model complexity and learning process to be adjusted to the dataset. The resulting train-test gap of XGBoost (0.070) was also smaller than that of Random Forest (0.163). In contrast, the lower performance of SVM-RBF may be related to its sensitivity to C and gamma in defining nonlinear decision boundaries. For the heterogeneous feature distributions in this dataset, the selected C and gamma configuration may not have separated the landslide and non-landslide samples as effectively as the tree-based models. Despite hyperparameter optimization, the selected SVM-RBF configuration did not represent the observed geospatial patterns as effectively as the tree-based models.

Differences from previous studies indicate that no algorithm consistently performs best across regions and datasets. Le et al. [12] and Halder et al. [13] reported high XGBoost performance, whereas Yao et al. [14] and Zhao et al. [15] reported better Random Forest performance. Such differences may arise from regional characteristics, feature composition, spatial resolution, sampling strategies, and algorithm configurations. In this study, XGBoost achieved the highest performance under the applied dataset and evaluation scheme, with McNemar's test showing significant differences from Random Forest and SVM-RBF. This finding supports the suitability of XGBoost for the Wonosobo dataset rather than its universal superiority.

Rainfall had the highest SHAP contribution, consistent with the mechanisms of water infiltration, increased pore-water pressure, reduced effective stress, and decreased soil shear strength [14]. The SHAP dependence plot further indicates that high rainfall contributes more positively to landslide predictions under steeper slopes, suggesting that the model captured their interaction. Dist_roads, slope, and elevation also showed substantial contributions, highlighting the combined roles of anthropogenic and topographic factors. Mean absolute SHAP values represent the relative magnitude of feature contributions, whereas their direction is reflected in the SHAP value distribution.

Landuse had the lowest mean absolute SHAP value (0.031), indicating a relatively smaller contribution to model predictions in this dataset rather than an absence of physical influence on landslides. This lower contribution may be related to the resampling of 10 m land-use data to 30 m, which may reduce fine-scale land-use information, as well as the stronger predictive contribution of rainfall and topographic factors such as slope and elevation. Differences from studies reporting landuse as an important factor may therefore reflect variations in spatial resolution, land-use distribution, sampling, study area, and feature composition.

The Jenks classification produced predominantly Low (49.2%) and High (44.4%) classes, while Moderate accounted for 6.4%, consistent with the concentration of predicted probabilities in low and high ranges. However, the thresholds of 0.326 and 0.835 are data-driven and should not be directly generalized to new datasets or regions. Key limitations include the relatively small sample size (642), 30 m spatial resolution, limited detailed soil and geological variables, and potential spatial autocorrelation. Although stratified cross-validation preserves class proportions, it does not specifically address spatial dependence. Future studies should consider spatial cross-validation, additional soil and geological factors, and validation across different areas or periods. SHAP should also be interpreted as a model explanation method rather than evidence of causal relationships.

4. CONCLUSION

This study developed and evaluated geospatial data-based landslide susceptibility classification models for Wonosobo Regency using Random Forest, Support Vector Machine with a radial basis function kernel (SVM-RBF), and XGBoost. Among the evaluated models, XGBoost provided the best predictive performance for the dataset, spatial predictors, and evaluation scheme used in this study, achieving a test accuracy of 93.02%, Cohen's Kappa of 0.861, and an AUC-ROC of 0.974. McNemar's test further indicated statistically significant differences between XGBoost and both Random Forest and SVM-RBF ($p < 0.05$). These findings indicate that, within the scope of this study, XGBoost was the most suitable model for classifying landslide susceptibility in Wonosobo Regency.

The results also identified rainfall as the most influential predictor, followed by distance to roads and slope gradient. The interaction between rainfall and slope indicates that sampled locations characterized by higher rainfall values and steeper terrain tend to receive higher predicted landslide susceptibility. Using the Jenks Natural Breaks classification, the predicted susceptibility values at the sampled locations were categorized into low, moderate, and high classes, accounting for 49.2%, 6.4%, and 44.4% of the sampled points, respectively. These proportions describe the distribution of the evaluated samples and should not be interpreted as the spatial distribution of susceptibility across Wonosobo Regency. The resulting sample-point susceptibility classification provides quantitative information on the model's predictions at the evaluated locations and may support further assessment of landslide susceptibility. However, the susceptibility classification at the sampled locations should be interpreted as a predictive assessment rather than a direct early warning system, because this study did not implement or validate real-time monitoring or warning mechanisms.

The findings should be interpreted considering the limitations of the available data and modeling framework. The analysis was based on 642 samples and 30-m spatial resolution data, while soil and geological variables were not included because of data availability constraints. Future research should expand the dataset using multitemporal observations and incorporate lithological and soil properties to improve the representation of landslide-controlling conditions. Field validation of sampled locations classified as highly susceptible is also needed to assess the practical reliability of the susceptibility classification. Further studies may evaluate spatial relationship-based models, such as Graph Neural Networks, as well as probability calibration and alternative classification thresholds to examine the stability of susceptibility estimates and the representation of intermediate-susceptibility areas.

5. REFERENCES

- [1] S. Fidan, H. Tanyaş, A. Akbaş, L. Lombardo, D. N. Petley, and T. Görüm, "Understanding fatal landslides at global scales: a summary of topographic, climatic, and anthropogenic perspectives,"

- Nat. Hazards*, vol. 120, no. 7, pp. 6437–6455, 2024, doi: 10.1007/s11069-024-06487-3.
- [2] M. S. Lanto, D. N. Syam, A. Wahyuni, and A. Kusmiran, “Pemetaan Risiko Bencana Tanah Longsor Di Kabupaten Barru Menggunakan Metode Pembobotan Parameter Dan Inarisk Bnpb,” *J. Geosaintek*, vol. 8, no. 3, p. 250, 2022, doi: 10.12962/j25023659.v8i3.14758.
- [3] M. R. B. E. Prasetyo, “BNPB Butuh Kesepahaman Perkuat Mitigasi Bencana Tanah Longsor,” 2025. Accessed: May 09, 2026. [Online]. Available: <https://www.antaranews.com/berita/4874533/bnpb-butuh-kesepahaman-perkuat-mitigasi-bencana-tanah-longsor>
- [4] A. Sugiharto, “Integrasi Iman dan Ekologi : Analisis Ekoteologi Kristen Terhadap Banjir dan Tanah Longsor 2025,” Vol. 10, No. 1, 2026.
- [5] H. M. Darojati, “Analisis Kesesuaian Sebaran Permukiman Dalam Upaya Mitigasi Bencana Tanah Longsor (Studi di Desabinangun, Kecamatan Watumalang, Kabupaten Wonosobo),” Sekolah Tinggi Pertanahan Nasional Yogyakarta, 2024.
- [6] M. Fakhri and S. S. Jumadi, “Identifikasi Lokasi Potensial Untuk Relokasi Permukiman Rawan Bencana Longsor di Kabupaten Wonosobo,” 2023, [Online]. Available: <https://eprints.ums.ac.id/id/eprint/108037>
- [7] T. A. Ndofah and P. B. Santosa, “Evaluasi Penggunaan Lahan Mengacu pada Indeks Potensi Lahan dan Kesesuaiannya Terhadap Rencana Tata Ruang Wilayah di Kabupaten Wonosobo,” *JGISE J. Geospatial Inf. Sci. Eng.*, vol. 6, no. 2, p. 87, 2023, doi: 10.22146/jgise.91079.
- [8] “Infografis Kejadian Bencana Tahun 2024,” 2025.
- [9] P. K. Badapalli *et al.*, “Landslide Susceptibility Level Mapping in Kozhikode, Kerala, Using Machine Learning-Based Random Forest, Remote Sensing, and GIS Techniques,” *Land*, vol. 14, no. 7, pp. 1–25, 2025, doi: 10.3390/land14071453.
- [10] A. Mani, M. Kumari, and R. Badola, “Landslide hazard zonation (LHZ) mapping of Doon Valley using multi-criteria analysis method based on remote sensing and GIS techniques,” *Discov. Geosci.*, vol. 2, no. 1, 2024, doi: 10.1007/s44288-024-00044-y.
- [11] M. A. Haq *et al.*, “Analysis of environmental factors using AI and ML methods,” *Sci. Rep.*, vol. 12, no. 1, pp. 1–16, 2022, doi: 10.1038/s41598-022-16665-7.
- [12] X. H. Le, C. Choi, S. Eu, M. Yeon, and G. Lee, “Quantitative evaluation of uncertainty and interpretability in machine learning-based landslide susceptibility mapping through feature selection and explainable AI,” *Front. Environ. Sci.*, vol. 12, no. July, pp. 1–15, 2024, doi: 10.3389/fenvs.2024.1424988.
- [13] K. Halder *et al.*, “Improving landslide susceptibility prediction through ensemble recursive feature elimination and meta-learning framework,” *Sci. Rep.*, vol. 15, no. 1, pp. 1–20, 2025, doi: 10.1038/s41598-025-87587-3.
- [14] J. Yao, X. Yao, Z. Zhao, and X. Liu, “Performance comparison of landslide susceptibility mapping under multiple machine-learning based models considering InSAR deformation: a case study of the upper Jinsha River,” *Geomatics, Nat. Hazards Risk*, vol. 14, no. 1, p., 2023, doi: 10.1080/19475705.2023.2212833.
- [15] Z. Zhao, Z. yuan Liu, and C. Xu, “Slope Unit-Based Landslide Susceptibility Mapping Using Certainty Factor, Support Vector Machine, Random Forest, CF-SVM and CF-RF Models,” *Front. Earth Sci.*, vol. 9, no. March, pp. 1–16, 2021, doi: 10.3389/feart.2021.589630.
- [16] K. L. Du, R. Zhang, B. Jiang, J. Zeng, and J. Lu, “Foundations and Innovations in Data Fusion and Ensemble Learning for Effective Consensus,” *Mathematics*, vol. 13, no. 4, pp. 1–49, 2025, doi: 10.3390/math13040587.
- [17] K. L. Du, B. Jiang, J. Lu, J. Hua, and M. N. S. Swamy, “Exploring Kernel Machines and Support Vector Machines: Principles, Techniques, and Future Directions,” *Mathematics*, vol. 12, no. 24, pp. 1–58, 2024, doi: 10.3390/math12243935.
- [18] Z. Rais, S. Muhammad Fahmuddin, Saida, and A. T. Utomo, “Implementation of Machine Learning Algorithm with Extreme Gradient Boosting (XGBoost) Method in Hypertension Level Classification,” *J. Appl. Sci. Eng. Technol. Educ.*, vol. 7, no. 1, pp. 126–136, 2025, doi: 10.35877/454RI.asci4191.
- [19] N. Widyaningrum, A. Fudholi, and E. Prawita Setyowati, “The Effect of Altitude Against Total Phenolic and Epigallocatechin Gallate (EGCG) Content in Green Tea Leaves,” *Proc. IConSSE FSM SWCU*, pp. 34–40, 2015.

- [20] A. Addis, “GIS-Based Landslide Susceptibility Mapping Using Frequency Ratio and Shannon Entropy Models in Dejen District, Northwestern Ethiopia,” *J. Eng. (United Kingdom)*, vol. 2023, pp. 115–128, 2023, doi: 10.1155/2023/1062388.
- [21] P. Pérez-Cutillas, A. Pérez-Navarro, C. Conesa-García, D. A. Zema, and J. P. Amado-Álvarez, “What is going on within google earth engine? A systematic review and meta-analysis,” *Remote Sens. Appl. Soc. Environ.*, vol. 29, no. September 2022, p. 100907, 2023, doi: 10.1016/j.rsase.2022.100907.
- [22] A. Velastegui-Montoya, N. Montalván-Burbano, P. Carrión-Mero, H. Rivera-Torres, L. Sadeck, and M. Adami, “Google Earth Engine: A Global Analysis and Future Trends,” *Remote Sens.*, vol. 15, no. 14, 2023, doi: 10.3390/rs15143675.
- [23] B. Alshehri, Z. Zhang, and X. Liu, “A Review of Google Earth Engine for Land Use and Land Cover Change Analysis: Trends, Applications, and Challenges,” *ISPRS Int. J. Geo-Information*, vol. 14, no. 11, pp. 1–26, 2025, doi: 10.3390/ijgi14110416.
- [24] M. Vizzari, “Object-Based Land Cover Classification in Google Earth Engine,” *Remote Sens.*, vol. 14, pp. 1–19, 2022.
- [25] K. Maharana, S. Mondal, and B. Nemade, “A review: Data pre-processing and data augmentation techniques,” *Glob. Transitions Proc.*, vol. 3, no. 1, pp. 91–99, 2022, doi: 10.1016/j.gltp.2022.04.020.
- [26] Q. Zhao and C. Poudyal, “Credibility theory based on winsorizing,” *Eur. Actuar. J.*, vol. 14, no. 3, pp. 777–807, 2024, doi: 10.1007/s13385-024-00391-7.
- [27] Y. W. Rabby, Y. Li, and H. Hilafu, “An objective absence data sampling method for landslide susceptibility mapping,” *Sci. Rep.*, vol. 13, no. 1, pp. 1–15, 2023, doi: 10.1038/s41598-023-28991-5.
- [28] J. Kotzé, J. Le Roux, and J. van Tol, “Investigating Soil Properties at Landslide Locations in the Eastern Cape Province, South Africa,” *GeoHazards*, vol. 6, no. 4, pp. 1–16, 2025, doi: 10.3390/geohazards6040068.
- [29] R. Gui *et al.*, “Digital Elevation Model-Driven River Channel Boundary Monitoring Using the Natural Breaks (Jenks) Method,” *Remote Sens.*, vol. 17, no. 6, pp. 1–24, 2025, doi: 10.3390/rs17061092.
- [30] M. Azad, T. H. Nehal, and M. Moshkov, *A novel ensemble learning method using majority based voting of multiple selective decision trees*, vol. 107, no. 1. Springer Vienna, 2025. doi: 10.1007/s00607-024-01394-8.
- [31] A. Roy and S. Chakraborty, “Support vector machine in structural reliability analysis: A review,” *Reliab. Eng. Syst. Saf.*, vol. 233, no. April, 2023, doi: 10.1016/j.res.2023.109126.
- [32] M. Wiens, A. Verone-Boyle, N. Henscheid, J. T. Podichetty, and J. Burton, “A Tutorial and Use Case Example of the eXtreme Gradient Boosting (XGBoost) Artificial Intelligence Algorithm for Drug Development Applications,” *Clin. Transl. Sci.*, vol. 18, no. 3, 2025, doi: 10.1111/cts.70172.
- [33] M. Castejón-Limas, L. Fernández-Robles, H. Alaiz-Moretón, J. Cifuentes-Rodriguez, and C. Fernández-Llamas, “A Framework for the Optimization of Complex Cyber-Physical Systems via Directed Acyclic Graph,” *Sensors*, vol. 22, no. 4, 2022, doi: 10.3390/s22041490.
- [34] K. M. Sujon, R. Hassan, K. Choi, and M. A. Samad, “Accuracy, precision, recall, f1-score, or MCC? empirical evidence from advanced statistics, ML, and XAI for evaluating business predictive models,” *J. Big Data*, vol. 12, no. 1, 2025, doi: 10.1186/s40537-025-01313-4.
- [35] D. Chicco and G. Jurman, “The Matthews correlation coefficient (MCC) should replace the ROC AUC as the standard metric for assessing binary classification,” *BioData Min.*, vol. 16, no. 1, pp. 1–23, 2023, doi: 10.1186/s13040-023-00322-4.
- [36] D. Khan, W. Akram, and S. Ullah, “Enhancing landslide susceptibility predictions with XGBoost and SHAP: a data-driven explainable AI method,” *Geocarto Int.*, vol. 40, no. 1, p., 2025, doi: 10.1080/10106049.2025.2514725.