Environmental Earth Sciences (2026) 85:101
https://doi.org/10.1007/s12665-026-12822-7

ORIGINAL ARTICLE

Optimization of various machine learning concepts to evaluate landslide susceptibility: XGBoost, k-NN and MLP using PSO algorithm

Hazem Ghassan Abdo[1,2] · Sahar Mohammed Richi[3] · Bilel Zerouali[4] · Saeed Alqadhi[5] · Okan Mert Katipoğlu[6] · Pankaj Prasad[7] · Hasan Arman[8] · Jasem A Albanai[9] · Javed Mallick[5]

Received: 20 May 2025 / Accepted: 17 January 2026

Abstract
Landslides significantly threaten natural and built environments, necessitating accurate prediction models for effective hazard mitigation. There is an urgent need to further improve the performance of machine learning algorithms in predicting landslide susceptibility by monitoring the impact of optimization algorithms on the performance of these models. This study evaluates the performance of various machine learning classifiers, including k-Nearest Neighbors (kNN), Multi-Layer Perceptron (MLP), and Extreme Gradient Boosting (XGBoost), for landslide susceptibility mapping. Additionally, Particle Swarm Optimization (PSO) is employed to enhance model performance by optimizing hyperparameters. Mountainous areas in the eastern Mediterranean (the northern Kabir River basin in western Syria) were identified as a result of the high frequency of landslide events over the past two decades. Nineteen factors causing landslides were identified, with no factor excluded, as a result of a multicollinearity test. The results indicate that XGBoost achieves the highest performance among traditional models. When integrated with PSO, the PSO-XGBoost model further improves classification performance, demonstrating its robustness in handling complex spatial patterns. Feature importance analysis using SHAP confirms slope as the dominant factor, followed by TRI, rainfall, Aspect, TWI, and curvature, highlighting the role of topography and hydrology in landslide occurrence. Moderate lithology, NDVI, and LULC contributions and lower importance of Flow Accumulation and Soil Depth suggest complex environmental interactions. Model predictions show varying susceptibility distributions. PSO-MLP assigns the highest very high susceptibility (44.09%), while PSO-XGBoost provides a balanced classification (31.13%). The PSO-XGBoost model demonstrates superior predictive capability, offering reliable landslide susceptibility maps for disaster risk management and land-use planning.

Keywords Optimization; landslide susceptibility; XGBoost · K-NN · MLP · PSO

Introduction

A landslide is classified as a morphodynamic activity of slope materials formed by weathering processes, consisting of rock, soil, and debris, under the influence of various driving factors, particularly gravity, seismic and volcanic activity, and rainstorms (Wei et al. 2021, 2024; Dey and Das 2025; Tanoumand et al. 2025; Hallal et al. 2024; Dai et al. 2025). In addition to the acceleration of human activities, including uncontrolled urbanization, road construction, intensive agriculture, slope modification, and mining, landslides have devastating spatial consequences for human lives, infrastructure, economic activity, transportation, and ecological diversity, while also significantly disrupting the quality of services provided (Susena et al. 2025; Liu et al. 2024a, b; Hou et al. 2025; Badreldin et al. 2025; Tebbouche et al. 2022). The latest statistics indicate that approximately 3.8 $km^2$of land worldwide is threatened by landslide events. This means that more than 290 million people are at risk of landslides (Aslam et al. 2022). A study Froude and Petley (2018) indicates that more than 180,000 people have been killed by more than 37,000 landslide events worldwide.

Landslides, particularly those of the rocky type, represent a persistent geomorphological threat to mountainous and plateau structures in the Eastern Mediterranean, particularly the Syrian Coastal Mountains, the eastern and western

Extended author information available on the last page of the article

Published online: 11 February 2026

Lebanon mountain ranges, the Qalamoun mountain ranges, and the northern Jordanian mountains (Yang et al. 2024; Al-Shabeeb et al. 2022; Awawdeh et al. 2018). Furthermore, landslide events have increased in the Eastern Mediterranean following the massive seismic activity that occurred on February 6, 2023 (Karakas et al. 2024).

Landslides in river basins in the Eastern Mediterranean cause thousands of casualties annually, along with significant destruction of housing and infrastructure, and disruption to sustainable development (Ahmad et al. 2017; Al-Rawabdeh et al. 2024; Tekin et al. 2024; Segoni et al. 2024). In this regard, preventing landslides is impossible, and therefore this risk must be managed through a precise understanding of their trends and work to predict their occurrences. Furthermore, the importance of this study lies in its aim to improve landslide maps in an area that has been experiencing significant seismic activity for more than three years on the Syrian-Turkish border (Karakas et al. 2024). Therefore, there is an urgent need to develop new models that predict landslide susceptibility, particularly in the Eastern Mediterranean region.

The literature on improving landslide susceptibility maps using machine learning models indicates significant research efforts in this area (Singh et al. 2024; Meng et al. 2024; Ajin et al. 2024). However, a thorough review of this literature revealed a complete lack of improvements in the performance of algorithms such as Extreme gradient boost (XGBoost), Multilayer Perceptron (MLP) and k-nearest neighbors (kNN) in predicting landslide susceptibility. The use of optimization algorithms, particularly the Particle Swarm Optimization (PSO) algorithm, is considered one of the most important methods applied in improving the accuracy of landslide susceptibility maps.

The key novelty of this study lies in its integration of PSO with multiple machine learning classifiers—specifically XGBoost, MLP, and kNN—to systematically enhance hyperparameter tuning for landslide susceptibility mapping. While previous research has applied individual machine learning models or optimization techniques separately, this study uniquely conducts a comparative performance evaluation across several classifiers both with and without PSO, allowing for a comprehensive assessment of the optimization impact. The application to a highly landslide-prone mountainous region in the eastern Mediterranean, an area with limited prior modeling efforts, further strengthens the novelty by providing the first PSO-enhanced susceptibility maps for this specific geographic context.

The ultimate goal of this evaluation is to improve the performance of XGBoost, MLP and kNN algorithms in producing landslide susceptibility maps using the optimization algorithm of PSO. However, the robustness of the models proposed in this study is achieved through the large number of landslide factors used, in addition to the reliable model construction method. These models propose three high-resolution landslide susceptibility maps in a region suffering from significant data and spatial management constraints. Thus, they provide a sophisticated perspective that helps decision-makers manage landslide risk in the Eastern Mediterranean region.

Materials and methods

Study area

The northern Kabir River basin is located in the western part of Syria between Latitudes 34° 59′ – 34° 54′ North and Longitudes 36° 22′ – 35° 44′ East, with a total watershed area of 1,081 km^2 (Fig. 1). The geomorphological features vary from plain sedimentary in the west and plateau and mountainous in the middle and east of the basin. The dry and hot summer and the wet and cold winter are the main climatic features in the basin, which is subject to the Mediterranean climate. The agricultural economic activity is the basis of the economic life of the population in the research basin.

Landslides inventory map

Landslide inventory maps play a vital role in landslide susceptibility studies by representing the locations and types of landslides in a region and helping predict future occurrences (Guzzetti et al. 1999, 2012). In this study, the landslide inventory map was developed using data obtained from historical records from the remote sensing directory in Lattakia governorate. Although the original data were accurate, they were not entirely sufficient. Therefore, they were supplemented through fieldwork and interpretation of high-resolution aerial photographs to ensure greater completeness and accuracy. A total of 1,548 landslide points were identified within the study area, along with an equal number of non-landslide points to create a balanced dataset (Fig. 1). In the current analysis, non-landslide locations were selected to match the number of landslide occurrences (1,548 points) using minimize spatial sampling bias (Jianget al. 2025).

Specifically, we applied a spatial filtering approach that ensures a minimum separation distance of 800 m between landslide and non-landslide points, consistent with recent studies that demonstrate how selecting non-landslide samples outside a buffer zone can reduce spatial sampling bias. Non-landslide points were distributed more homogeneously across the study area while avoiding areas with similar

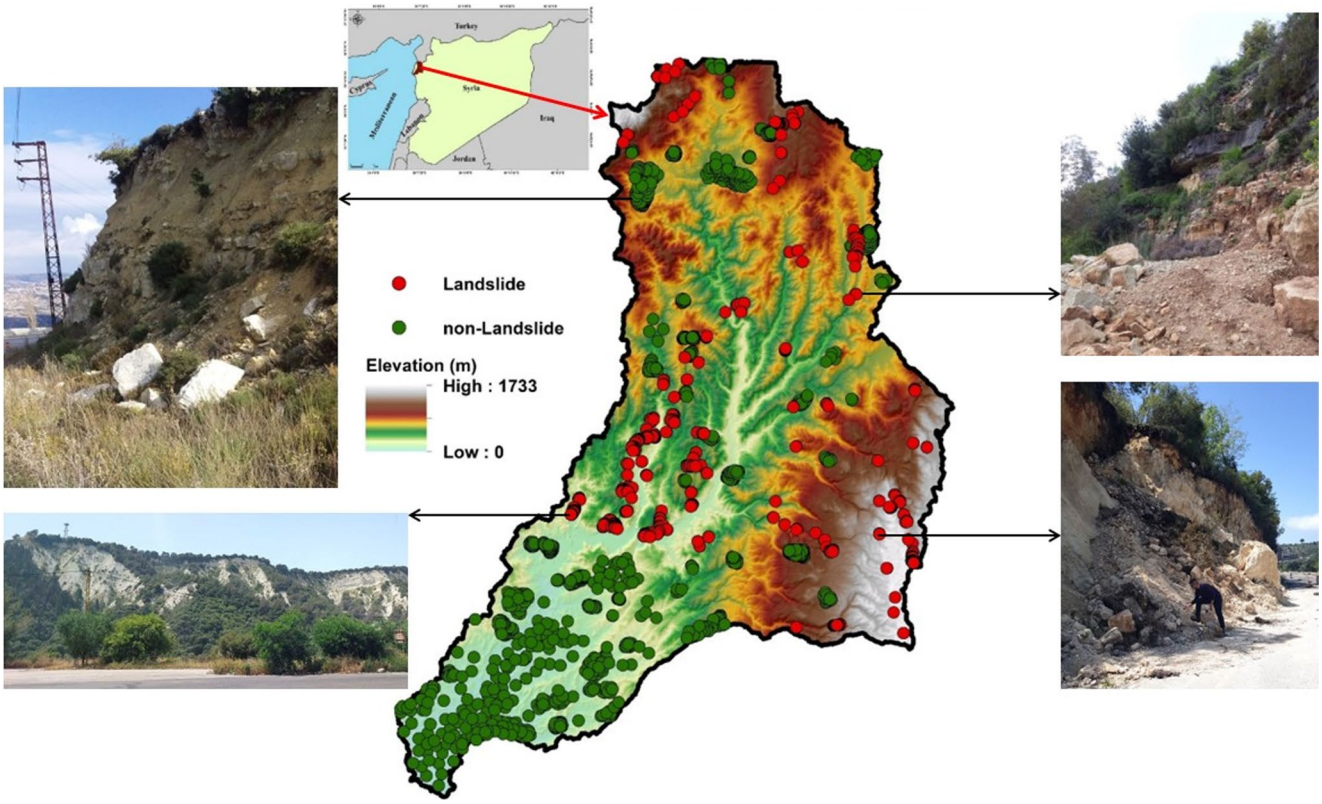

Fig. 1 Study area location

environmental conditions to known landslide sites, in order to prevent data leakage and inflated model performance. The dataset was initially divided into two subsets, with 70% allocated for training and 30% reserved for independent testing. The 70:30 ratio for splitting the dataset into training and testing subsets was chosen because it represents one of the most widely adopted and empirically validated practices in machine learning and geospatial modeling studies, including landslide susceptibility mapping. This proportion provides a good balance between ensuring that the training subset is large enough to capture the underlying data distribution for robust model learning, while still retaining a sufficiently large and independent testing subset for reliable model evaluation (Pawluszek-Filipiak and Borkowski 2020; Halder et al. 2025a, b, c).

To enhance the robustness and generalizability of the model performance estimates, the training subset was further subjected to a 5-fold cross-validation procedure. This approach ensures that each instance in the training data is used for both training and validation, thereby reducing the risk of bias and mitigating the overly optimistic performance estimates that may arise from a single random split. The independent test set was only employed for the final evaluation of the trained models.

Data acquisition and pre-processing

In the present study, several geographical and environmental conditioning factors were utilized to estimate the susceptibility of landslides using the optimized XGB, k-NN, and MLP algorithms with the PSO algorithm. The dataset includes topography, hydrology, geology, climatic conditions and land use parameters from different sources.

Topographic variables, including slope orientation, slope degree, plan and profile curvature, elevation, and various topographic indices, namely Stream Power Index (SPI), Topographic Position Index (TPI), Terrain Ruggedness Index (TRI), Sediment Transport Index (STI), and Topographic Wetness Index (TWI), were derived from the ALOS 12.5-meter resolution Digital Elevation Model (DEM) obtained from the Alaska Satellite Facility (ASF). These indices were generated using standard GIS spatial analysis tools, where the DEM was first pre-processed to fill sinks and remove artifacts to ensure topographical and hydrological consistency.Vegetation cover was represented by the Normalized Difference Vegetation Index (NDVI), extracted from Landsat-8 OLI satellite imagery. The imagery was pre-processed by atmospheric correction and cloud masking to ensure data quality. In this study, NDVI values were

computed as seasonal averages over the winter and spring months, which are the wet seasons in the study area. During these seasons, intense rainfall events are frequent, leading to elevated soil saturation and a higher incidence of landslides, as confirmed by field investigations. This seasonal averaging approach was selected to better represent vegetation cover conditions during the period of greatest slope instability, thereby capturing vegetation dynamics most relevant to landslide initiation.

Rainfall data were collected from the Syrian Meteorological Service (LMS), comprising long-term precipitation records. These data were spatially interpolated using the Inverse Distance Weighting (IDW) method to produce continuous rainfall surfaces over the study area.

Road network data were sourced from the Ministry of Public Works and Transportation (General Directorate of Roads and Buildings) and digitized into vector format to capture anthropogenic influence. Fault lines and lithological units were obtained from geological maps at a 1:50,000 scale issued by the Syrian Petroleum Authority. These maps were digitized and georeferenced to align with the DEM and other spatial layers. To address roads and faults spatially, the *Euclidean distance* index was used in a GIS environment, in line with the boundaries of the study area. Soil depth information was provided by the Ministry of Agriculture, based on field surveys and soil profile data. Land Use Land Cover (LULC) data were acquired from USGS dataset products, which were reclassified into relevant categories for susceptibility analysis.

All datasets were projected to a common coordinate system, resampled to match the spatial resolution of the DEM where applicable, and clipped to the study area boundary. Data cleaning involved checking for missing values, outliers, and ensuring spatial alignment. This comprehensive pre-processing workflow ensured the integrity and compatibility of multi-source data inputs for subsequent modeling stages (Table 1).

Table 1 Data sources, original Spatial resolutions, and preprocessing steps for the conditioning factors used in landslide susceptibility modeling

Conditioning factor	Data source	Original resolution	Preprocessing steps
Slope, Aspect, Curvature, Elevation, SPI, TPI, TRI, STI, TWI and flow accumulation	ALOS PALSAR DEM – Alaska Satellite Facility	12.5 m	Sink filling, slope/aspect/curvature derivation, index calculation, resampling
NDVI	Landsat-8 OLI	30 m	Atmospheric correction, cloud masking, NDVI calculation, resampling to 12.5 m
Rainfall	Syrian Meteorological Service	Point data	IDW interpolation to 12.5 m raster grid
Road Network	Ministry of Public Works & Transportation	Vector (line)	Digitization, rasterization to 12.5 m resolution
Fault Lines, Lithology	Syrian Petroleum Authority – Geological Maps	1:50,000 scale (vector)	Georeferencing, digitization, rasterization
Soil	Ministry of Agriculture	Vector/Tabular	Spatial join to DEM grid, rasterization to 12.5 m
LULC	Landsat-8 OLI	30 m	Reclassification, resampling to 12.5 m
Population Density	Census Data	100 m	Downscaling to 12.5 m using dasymetric mapping techniques

Analysis process

The analysis involved several systematic steps to ensure robust landslide susceptibility modeling. Initially, all spatial datasets were pre-processed and standardized to a common coordinate system and resolution. In this study, all spatial datasets were processed and analyzed using the World Geodetic System 1984 (WGS 84) geographic coordinate system (EPSG:4326), which is the most commonly adopted reference system in geospatial and remote sensing applications. For detailed spatial analyses and to minimize distortions in distance and area measurements, the data were projected into the Universal Transverse Mercator (UTM) Zone 37 N (EPSG:32637), which is the standard projected coordinate system covering western Syria, including the Lattakia governorate where the present study was conducted. This ensured consistency, spatial accuracy, and compatibility across all input datasets and derived products.

To avoid internal correlation between the influencing factors and overfitting, the correlation coefficient and multicollinearity values were calculated. Three machine learning classifiers, kNN, MLP, and XGBoost, were implemented to develop predictive models. Hyperparameter optimization was performed using PSO to enhance model performance. Model training and testing were carried out with a 70:30 split of the dataset, ensuring unbiased evaluation. Model accuracy, Area Under the Curve (AUC), and other performance metrics were calculated to compare classifier effectiveness. Additionally, feature importance was analyzed using SHAP (SHapley Additive exPlanations) values to interpret the contribution of each factor. The entire workflow was executed

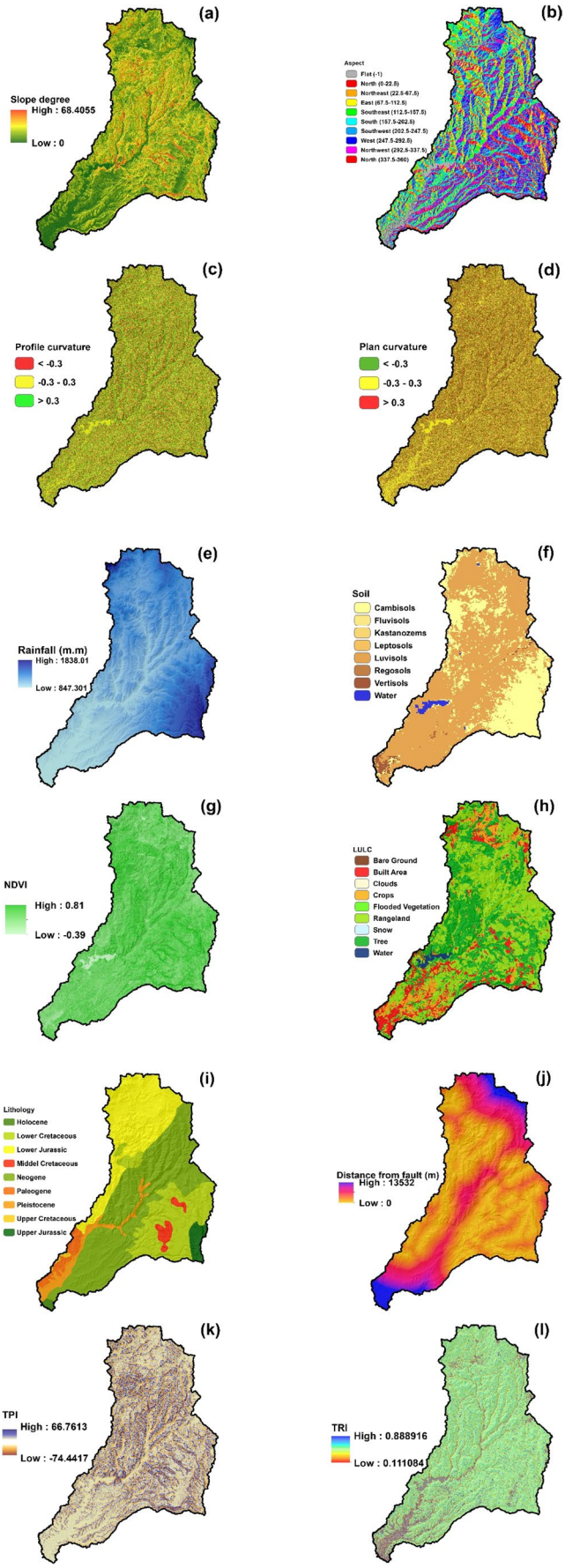

Fig. 2 Thematic layers of landslides causative factors (**a**) slope, (**b**) aspect, (**c**) profile curvature, (**d**) plan curvature, (**e**) rainfall, (**f**) soil type, (**g**) NDVI, (**h**) LULC, (**i**) lithology, (**j**) distance to faults, (**k**) TPI, (**l**) TRI, (**m**) TWI, (**n**) flow accumulation, (**o**) distance to rivers, (**p**) STI, (**q**) SPI, (**r**) distance to roads, (**s**) population

using the Python programming language, employing libraries such as scikit-learn, XGBoost, TensorFlow/Keras, and PySwarms for optimization. This comprehensive approach ensured rigorous model development, evaluation, and interpretability.

To ensure reproducibility and transparency, the software environments and tools employed in this study are specified as follows. All spatial data preprocessing, conditioning factor extraction, and map visualization were conducted using ArcGIS 10.3. The machine learning analyses were implemented in Python 3.11.4, with dedicated libraries applied according to specific tasks: scikit-learn for model training, evaluation, and preprocessing; xgboost for implementing the XGBoost algorithm; pyswarm for PSO; numpy and pandas for numerical computation and data handling; statsmodels for statistical analysis; and matplotlib and seaborn for result visualization.

Landslides predisposing factors

Slope

One of the most critical factors in landslide susceptibility is that steeper slopes are more prone to failure due to gravitational forces and weaker shear strength. Steep slope triggers landslides by affecting the change of stress field, which is related to the infiltration of the slope and the landslide (Hong et al. 2016, 2019b). Figure 2a shows the slope values in the study basin.

Aspect

Aspect is an effective and important topographic factor in the occurrence of landslides by affecting rainfall infiltration and runoff, moisture retention, absorption of solar radiation, weathering conditions soil stability and vegetation change (Wang et al. 2019). Figure 2b presents the aspect factor in the study area.

Profile curvature

Profile curvature influences the speed of surface water runoff and soil displacement along the slope. Profile curvature is an effective topographic variable that determines landslide formation by regulating erosion and surface runoff and shows the geometric structure of the slope (Moosavi and Niazi 2016; Chen et al. 2021). The spatial distribution of the

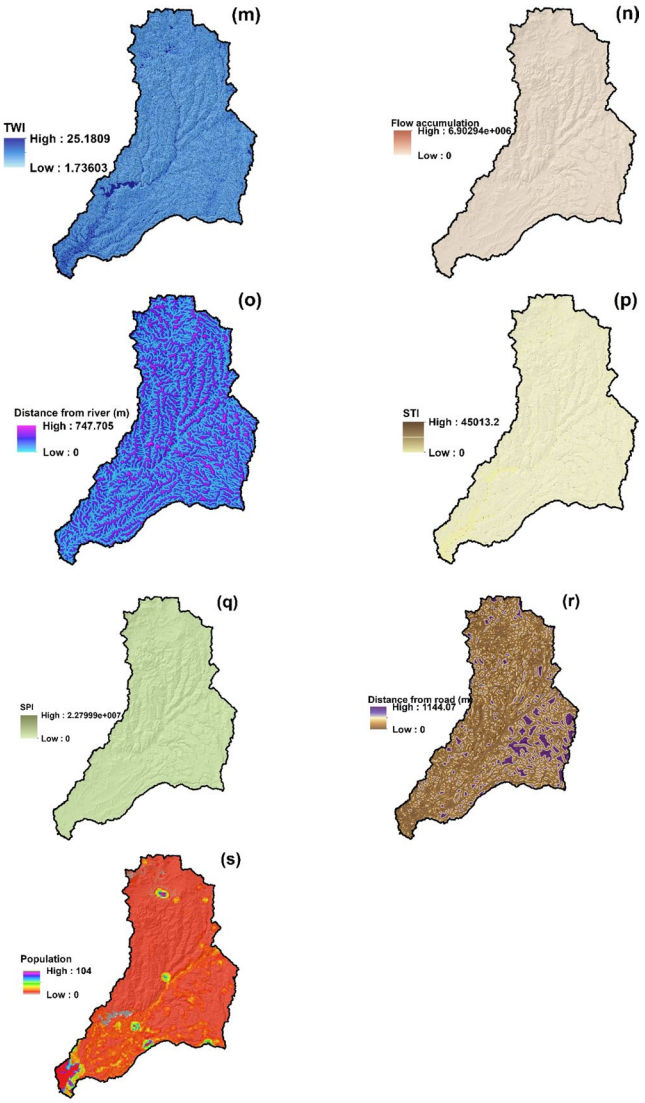

Fig. 2 (continued)

profile curvature values varies in the study area as shown in the Fig. 2c.

Plan curvature

Plan curvature is a parameter that should be taken into account in landslide formation by affecting the convergence or divergence of surface flow, water concentration and erosion intensity (Moosavi and Niazi 2016; Chen et al. 2021). Figure 2d shows the plan curvature values in the study basin.

Rainfall

Heavy precipitation is a primary landslide trigger, increasing pore water pressure and weakening soil shear strength. Short-term high-intensity rainfall will increase the pore water pressure of the land and deteriorate the slope stability. Thus, the risk of landslides will increase. (Sun vd., 2020b, 2021). Figure 2e shows the rainfall values in the study basin.

Soil types

Different soil types vary in shear strength, permeability, and water infiltration capacity, affecting landslide occurrence (Francone 2022). Figure 2f presents the soil types factor in the study area.

NDVI

NDVI measures vegetation density, which helps stabilize slopes by reinforcing soil and reducing surface runoff. NDVI is an important environmental factor affecting the flow and erosion of the slope (Zhang et al. 2019). NDVI is a parameter expressing the vegetation characteristics in determining the susceptibility zonation (Yi et al. 2020). NDVI given by Eq. 1.

$$NDVI = (NIR - R)/(NIR + R) \quad (1)$$

here NIR and R denote the electromagnetic spectrum of the near-infrared and red bands of the satellite image, respectively (Fig. 2g).

LULC

Human-induced land-use changes, such as deforestation and urbanization, alter surface runoff and soil cohesion, affecting landslide susceptibility. LULC has an effective say in determining landslide risk because it has a significant effect on the soil's water-holding capacity, infiltration rate, surface flow and erosion resistance. In forested areas, the root systems of trees stabilize the soil and reduce the risk of landslides. However, when heavy rainfall occurs, the risk of landslides may increase as tree roots reach saturation. In urban areas, due to the abundance of concreting and asphalt surfaces, the rate of water naturally passing to the surface flow increases. Thus, landslide potential may increase in sloping areas (Rabby et al. 2022; Tyagi et al. 2023). The spatial distribution of the LULC varies in the study area as shown in the Fig. 2h.

Lithology

The composition and structure of rock formations play a vital role in determining slope stability (Shirzadi vd., 2018). Different lithology means that the physical and chemical structures of rocks are variable, which creates great effects on the slope. Generally, areas where low permeability rock structure is dominant have a higher landslide risk (Zhou et

al. 2021). The spatial distribution of the lithology varies in the study area as shown in the Fig. 2i.

Distance to faults

Distance to faults being near geological faults raises landslide risk due to seismic activity and ground displacement (Yi vd., 2020). The proximity and discontinuity of the region to the fault are directly proportional to the looseness of the soil and the landslide risk (Zhou et al. 2021). Figure 2j presents the distance to faults factor in the study area.

TPI

TPI is a topographic index used to express the difference in height between two cells. TPI identifies landforms such as ridges, valleys, and plains, which influence landslide occurrence (Alemayehu et al. 2018). High TPI values indicate ridges, while negative TPI values indicate valleys. Also, TPI values close to zero indicate steep or flat terrains. Figure 2k presents the TPI factor in the study area.

TRI

TRI is an indicator of the roughness of the terrain by evaluating the elevation differences in a region (Riley and Degloria 1999). In landslide-prone regions, rugged, steep slopes and regions with significant elevation fluctuations are seen to have a high risk of landslides (Al-Najjar et al. 2023). TRI given by Eq. 2.

$$TRI = Abs(max^2 - min^2) \tag{2}$$

where max and min denote the maximum and minimum elevation of the study area, respectively. The spatial distribution of the TRI values varies in the study area as shown in the Fig. 2l.

TWI

TWI represents soil moisture accumulation potential, influencing the likelihood of slope failure due to water saturation. TWI is an important parameter in determining how topography affects hydrological events. In areas with high TWI values, there is a higher potential for soil moisture content, meaning that the soil is likely to become saturated and slide (Sun et al. 2020c). TWI values are calculated by Eq. 3 below.

$$TWI = In(A/tan\beta) \tag{3}$$

here A denotes flow accumulation and β indicates the slope gradient at each pixel of the image. The spatial distribution of the TWI values varies in the study area as shown in the Fig. 2m.

Flow accumulation

Flow accumulation identifies areas where water collects, leading to soil saturation and potential landslide initiation. Flow accumulation is a hydrotopographic parameter that shows the direction of water flow and the amount of flow collected at a certain point (Wu et al. 2024). It is determined by the Digital Elevation Model (DEM) in the GIS environment. It is calculated using and helps to analyze how water accumulates on sloping lands. Landslide risk is high with high flow accumulation on slopes with loose material (Yuan et al. 2022). The spatial distribution of the flow accumulation values varies in the study area as shown in the Fig. 2n.

Distance to river

The closer a slope is to a river, the more vulnerable it is to erosion and undercutting, increasing landslide susceptibility. The stability of slopes is significantly influenced by fluvial erosion and cutting. Furthermore, fluctuations in water levels significantly impact the groundwater environment of landslide-prone areas (Zeng et al. 2023a, b). Figure 2o presents the distance to river factor in the study area.

STI

Assesses sediment transport potential, which impacts erosion and landslide susceptibility. Large STI values cause the erosion and transportation capacity of the water flow to be high. Thus, the landslide potential of the slope adjacent to the river is high (Moore and Wilson 1992; Pradhan and Kim 2014). STI given by Eq. 4.

$$STI = \left(\frac{A}{22.13}\right)^{0.6}\left(\frac{sin\beta}{0.0896}\right)^{1.3} \tag{4}$$

The spatial distribution of the STI values varies in the study area as shown in the Fig. 2p.

SPI

SPI, measures the erosive power of flowing water, contributing to slope instability. The larger the SPI values, the greater the erosion and transportation capacity of the water flow. Thus, the landslide risk of the slope adjacent to the river increases (Moore and Wilson 1992; Ali et al. 2021). SPI values are calculated by Eq. 5 below.

$$SPI = A \times tan\beta \tag{5}$$

The spatial distribution of the SPI values varies in the study area as shown in the Fig. 2q.

Distance to roads

Infrastructure development disrupts natural drainage patterns and slope stability, increasing landslide risks in poorly managed areas (Pham et al. 2017). Figure 2r presents the distance to roads factor in the study area.

Population

Human activities, including land modification and construction, can exacerbate landslide susceptibility by altering the landscape (Yu et al. 2023; Zeng et al. 2023a, b; Taher et al. 2025). The spatial distribution of the population values varies in the study area as shown in the Fig. 2s.

Multicollinearity test

The identification of relevant factors influencing landslide occurrence is fundamental to improving the accuracy of landslide susceptibility modeling. Previous research has applied various techniques for variable selection, such as Principal Component Analysis (PCA), multi-collinearity analysis, information gain, and the Boruta algorithm. In this study, multi-collinearity analysis was employed to determine the most significant predictors for landslide susceptibility mapping. This statistical method is widely used to remove highly correlated variables, thereby enhancing the stability and precision of predictive models. To ensure the robustness of the analysis, the Variance Inflation Factor (VIF) and Tolerance (TOL) were used as diagnostic indicators, with selection criteria of VIF values less than 10 and TOL values greater than 0.5 (Selamat et al. 2025; Bravo-López et al. 2023). Variables that exceeded these thresholds were considered to exhibit problematic multicollinearity and were excluded from the final model. The VIF was calculated using the standard formulation (Eq. 6).

$$VIF = \frac{1}{Tolerance} \tag{6}$$

ML algorithms

Extreme gradient boost (XGBoost)

XGBoost is a tree-based ensemble learning approach that is a member of the Gradient Boosting Machines (GBM) family. In turn, XGBoost combines its weak learners, which are decision trees, to form a strong tree structure. Each new learner aims to develop an effective ensemble structure by correcting the errors produced by its predecessors. XGBoost has important advantages such as providing high accuracy in solving engineering problems and being resistant to overfitting problems (Chen and Guestrin 2016).

Multilayer perceptron (MLP)

MLP is trained based on the principles of forward and backward propagation. First, the data entering from the input layer is sent to the neurons in the hidden layer. Each neuron is weighted with a certain weight by multiplying the inputs separately, adding bias, and finally passing to the activation function. While this is in the hidden layer, all of the neurons are repeated to reach the output layer and the final prediction is made. Then, backpropagation is applied and the performance and weight are updated. This process is repeated many times to learn the network and minimize the model error rate (Sahana et al. 2022; Meshram et al. 2022).

k-nearest neighbors (kNN)

kNN is a distance-based machine learning (ML) approach used for effective and fast solutions to both regression and classification problems. This approach performs the learning process by using the training data as the "base" for the predicted values. The basic working principle is that neighbors are selected from the base and certain are ranked according to the degree of similarity. To determine the degree of similarity (i.e. Manhattan Distance) (Bansal et al. 2022; Iheanetu and Obileke 2024). d_j given by Eq. 7.

$$d_j = \sum_{k=1}^{n} W_k \left| x_{\text{train},j,k} - x_{\text{test},k} \right| \tag{7}$$

where d_j is the distance between the i_{th} training and test data, W_k is the weight of the j_{th} attribute and training data, and the attribute values of the training data and test are x_{train} and x_{test}, respectively. where j and k denote the training data and the test attribute, respectively, and n denotes the number of attributes.

PSO-based optimization algorithms

PSO is a metaheuristic algorithm inspired by the behavior of fish and bird flocks in nature. Each particle in PSO represents a possible solution. Each particle moves by following the current optimum particles. PSO is based on velocity and position updates. It is based on finding a high-quality solution by searching in space for a swarm of n particles and updating their positions and velocities (Sahab et al. 2013; Lu et al. 2018). The position and velocity of particle i are indicated $xi=(x_{i1},x_{i2},\ldots,x_{im})$ and $vi=(v_{i1},v_{i2},\ldots,v_{im})$. To regulate the velocities and direct the swarm, PSO defines two

positions: the personal-best position, defined as pbest and the global-best position, defined as gbest. The computation of velocity and position was performed using the Eq. 8.

$$v_i^{t+1} = \omega\, v_i^t + c_1 r_1 \left(p_{\text{best},i}^t - x_i^t\right) + c_2 r_2 \left(g_{\text{best}}^t - x_i^t\right) \qquad x_i^{t+1} = x_i^t + v_i^{t+1} \tag{8}$$

where c_1=cognitive factor, c_2=social factor, ω=inertia weight, r_1,r_2=real numbers, and t=current generation (Huang et al. 2016).

Hyperparameters ranges and PSO configuration

To enhance clarity and ensure full reproducibility of the modeling process, the search ranges of the optimized hyperparameters for each algorithm, along with the configuration of the PSO procedure, are summarized in Table 2. For the kNN model, the optimized parameters included: *n_neighbors* (1–20), *weights* (uniform/distance), *metric* (Euclidean/Manhattan), *algorithm* (ball_tree/kd_tree/brute), *leaf_size* (1–50), *p* (1–2), and *n_jobs* (parallelization option). For the XGBoost model, the hyperparameters considered were: *n_estimators* (50–500), *max_depth* (1–10), *learning_rate* (0.01–0.3), *gamma* (0–5), *min_child_weight* (1–10), *subsample* (0.5–1.0), and *colsample_bytree* (0.5–1.0). For the MLP, the ranges were: *hidden_layer_sizes* (5–100), *activation* (identity, logistic, tanh, relu), *solver* (sgd/adam), *alpha* (0.0001–1), *learning_rate_init* (0.0001–0.1), and *max_iter* (100–500). The PSO algorithm was implemented using the *pyswarm* library, with the following configuration: swarm size=15, number of iterations=10, inertia weight (*w*)≈0.7298, cognitive coefficient (*c1*)=1.4962, and social coefficient (*c2*)=1.4962.

Table 2 Hyperparameters ranges and PSO configuration

Model	Hyperparameter	Search Range/Options
kNN	n_neighbors	1–20
	weights	uniform, distance
	metric	Euclidean, Manhattan
	algorithm	ball_tree, kd_tree, brute
	leaf_size	1–50
	p	1–2
	n_jobs	–1 (parallelization option)
XGBoost	n_estimators	50–500
	max_depth	1–10
	learning_rate	0.01–0.3
	gamma	0–5
	min_child_weight	1–10
	subsample	0.5–1.0
	colsample_bytree	0.5–1.0
MLP	hidden_layer_sizes	5–100
	activation	identity, logistic, tanh, relu
	solver	sgd, adam
	alpha	0.0001–1
	learning_rate_init	0.0001–0.1
	max_iter	100–500
PSO	Swarm size	15
	Iterations	10
	Inertia weight (*w*)	0.7298
	Cognitive coefficient (*c1*)	1.4962
	Social coefficient (*c2*)	1.4962

Performance assessment

The construction of the confusion matrix was predicated on multiple combinations of true and predicted values. For landslide predictions: TP (True Positive): Predicted landslide, actually occurred, TN (True Negative): Predicted no landslide, none occurred, FN (False Negative): Predicted no landslide, but occurred, FP (False Positive): Predicted landslide, but didn't occur. Accuracy (Acc), precision, recall, F-measure (Fang et al. 2020; Ji et al. 2020). ROC–AUC curve shows the relation between a true positive rate (sensitivity) and a false positive rate (1-specificity). Performance assessment formulated as below (Eqs. 9, 10, 11, 12, and 13).

$$Recall/Sensitivity = \frac{TP}{TP + FN} \tag{9}$$

$$Specificity = \frac{TN}{TN + FP} \tag{10}$$

$$Precision = \frac{TP}{TP + FP} \tag{11}$$

$$F1 = \frac{2TP}{2TP + FP + FN} \tag{12}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{13}$$

Results

Selection of landslide causative factors

Figure 3 presents the correlation matrix among the 19 conditioning factors used for landslide susceptibility modeling. The correlation coefficients range from −0.53 to 1.00, with the diagonal representing perfect self-correlation. Most factor pairs exhibit weak to moderate correlations, indicating limited redundancy among variables and supporting their inclusion in the modeling process. Strong positive correlations are observed between SPI and Flow accumulation (r=0.67), and Flow accumulation and Distance from river (r=0.62), reflecting logical hydrological

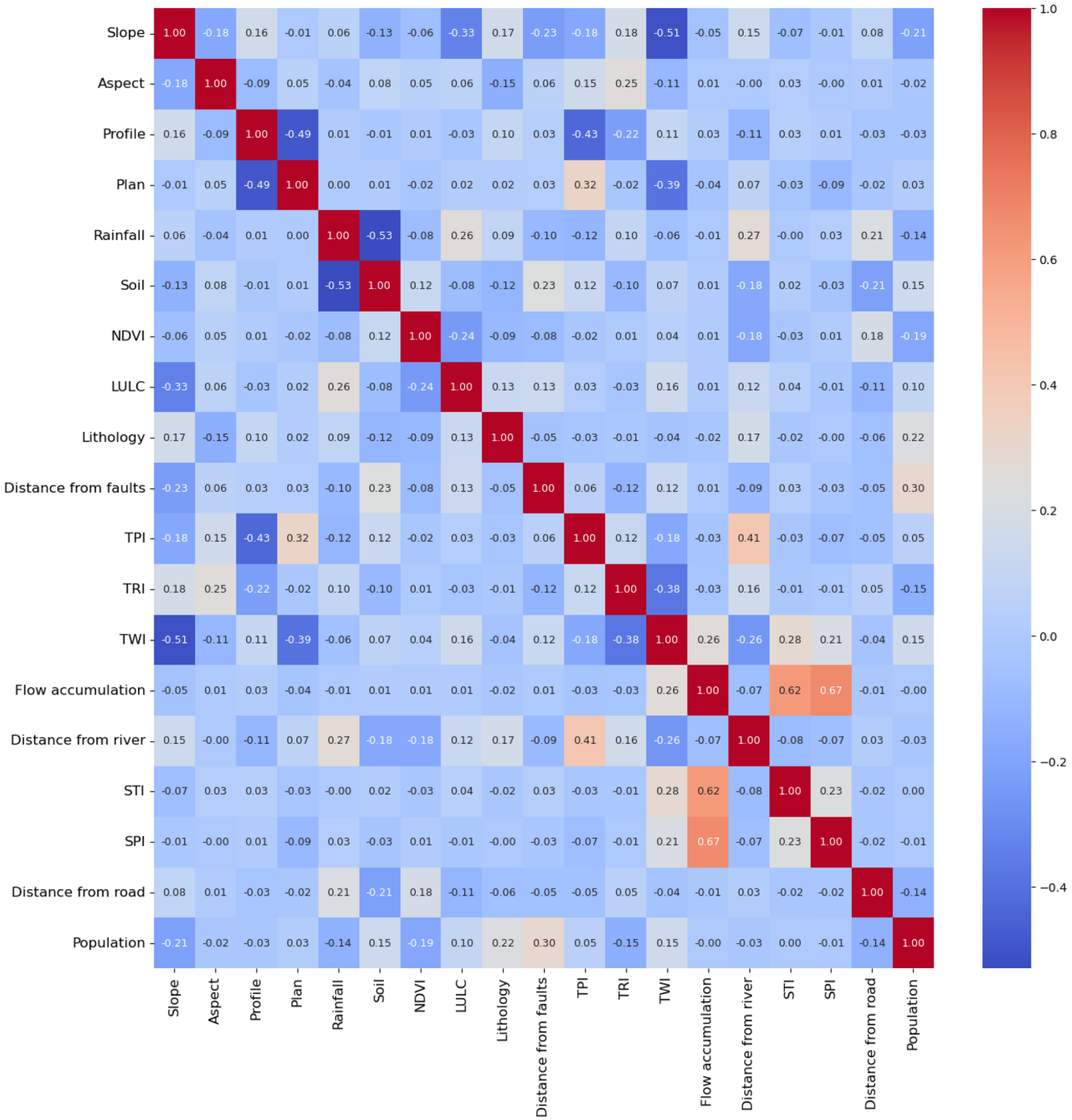

Fig. 3 Correlation matrix of the conditioning factors used in landslide susceptibility modeling

and geomorphological linkages. Conversely, notable negative correlations are found between Rainfall and Soil ($r = -0.53$), Profile curvature and Plan curvature ($r = -0.49$), and TPI and Profile curvature ($r = -0.43$), TWI and Slope ($r = -0.51$), highlighting potential terrain and soil–climate interactions. Population density exhibits low correlations with most environmental variables, reinforcing its independence as an anthropogenic factor. The absence of high multicollinearity among most factors minimizes overfitting risks and strengthens the robustness of the machine learning models applied in this study.

The VIF and TOL values, presented in Table 3, indicate that multicollinearity among the 19 conditioning factors used in the landslide susceptibility modeling is within acceptable limits. All VIF values are below the commonly used threshold of 5, suggesting the absence of severe

Table 3 Multicollinearity values for the 19 conditioning factors used in landslide susceptibility modeling

	Variable	VIF	TOL
1	Slope	2.072	0.48
2	Aspect	1.194	0.84
3	Profile	1.768	0.57
4	Plan	1.917	0.52
5	Rainfall	1.660	0.60
6	Soil	1.541	0.65
7	NDVI	1.221	0.82
8	LULC	1.408	0.71
9	Lithology	1.226	0.82
10	Distance from faults	1.215	0.82
11	TPI	1.793	0.56
12	TRI	1.510	0.66
13	TWI	2.599	0.38
14	Flow accumulation	3.165	0.32
15	Distance from river	1.602	0.62
16	STI	1.934	0.52
17	SPI	2.126	0.47
18	Distance from road	1.152	0.87
19	Population	1.308	0.76

multicollinearity issues that could distort model estimates. The highest VIF value is observed for Flow Accumulation (VIF = 3.165; Tolerance = 0.32), followed by TWI (VIF = 2.599; Tolerance = 0.38) and slope (VIF = 2.072; Tolerance = 0.48), indicating moderate correlation with other predictors. Conversely, variables such as distance from road (VIF = 1.152; Tolerance = 0.87), aspect (VIF = 1.194; Tolerance = 0.84), and distance from faults (VIF = 1.215; Tolerance = 0.82) show the lowest collinearity, reflecting their relative independence from other factors. Based on these results, and given the acceptable multicollinearity levels, all conditioning factors were retained in the modeling process to preserve the full spectrum of environmental and anthropogenic influences on landslide susceptibility. These results confirm that the selected predictors provide diverse and complementary information for the susceptibility modeling process, thereby enhancing the robustness and interpretability of the machine learning models.

The SHapley additive explanations analysis (SHAP)

The interaction between landslide occurrence and landslide triggering factors is not similar in each geographical area. For this, it is necessary to understand the role of landslide causative factors in a specific region. In the present study, the SHapley Additive exPlanations (SHAP) method has been used to calculate the importance of each landslide triggering factor. The presented SHAP plot in Fig. 4 illustrates the impact of different variables on the output of the landslide susceptibility classification model.

The slope is the most influential variable, with high values (red) strongly increasing landslide risk, while low slopes (blue) have a negative or neutral effect. This is because steep slopes reduce soil stability and increase gravitational force, thus promoting instability. Population has a significant impact, probably due to urbanization and anthropogenic landscape modifications, such as deforestation and construction, which make soils more vulnerable. The TPI also shows high variability, suggesting that crest or trough areas influence terrain stability. Aspect influences landslide susceptibility, as slope orientation plays a role in exposure to climatic conditions (rain, wind, sunlight). Distance to faults is also a critical factor, with areas close to fractured geological structures often being more unstable. Hydrological factors such as rainfall, TWI, and SPI have a significant impact. Precipitation increases soil saturation, reducing its cohesion and thus increasing the risk of landslides. Lithology plays a role in terrain stability, as certain rock and soil types are more prone to erosion and landslides.

The NDVI has a moderate influence, indicating that vegetation cover contributes to soil stabilization but is not a determining factor on its own. Other variables, such as distance to roads and distance to rivers, have a smaller impact, although infrastructure development can disrupt soil stability. Flow accumulation and the Soil show a reduced effect, suggesting that they act as secondary factors. In conclusion, the most influential factors are slope, rainfall, TPI, aspect, lithology, and distance to faults. These results confirm the importance of combining multiple environmental variables to improve landslide prediction.

Landslides susceptibility maps

The landslide susceptibility map is the final output of landslide assessment, which provides the spatial information of landslide intensity that can be used for landslide risk mitigation. After the model establishment and the significant accuracy from the testing dataset, each pixel of the research region has been extracted to predict the landslide susceptibility of the entire study area. This landslide susceptibility value of the region ranges between 0 and 1, where 0 indicates no chance of landslide events, and 1 represents a high chance of landslide occurrence. Based on the distribution and nature of the datasets, there are many classification methods, namely geometrical, manual, min-max normalization, standard deviation, equal intervals, and natural break, that have been used in natural hazard and resource mapping to classify the probability values (Kalantar et al. 2020; Lv et al. 2022; Prasad et al. 2022). In the current research, the natural break method has been employed according to the distribution of probability values achieved from applied

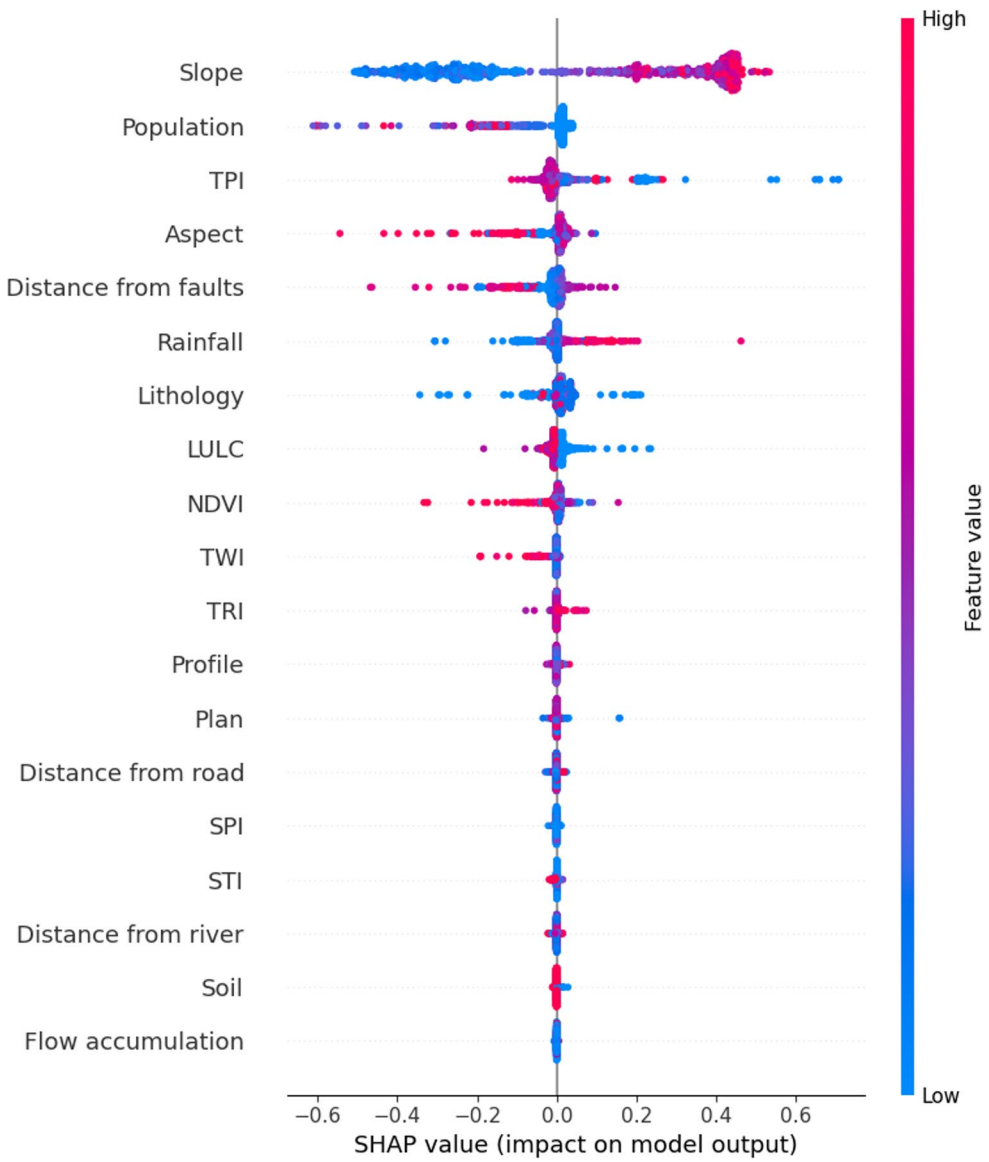

Fig. 4 Importance of landslides causative factors

machine learning models (Figs. 5 and 6). The landslide susceptibility maps were produced at a spatial resolution of 12.5 m, corresponding to the highest available resolution among the input data used, the 12.5 m ALOS DEM. This resolution was maintained throughout the processing and analysis stages to ensure spatial consistency between the different spatial layers and to achieve a more accurate representation of spatial variations in susceptibility.

The analysis of the results obtained for the classification of landslide susceptibility highlights notable differences between the models tested, both in terms of the spatial distribution of classes and in the proportion of risk areas. For the kNN model, the "Very Low" class occupies a significant area of 416.35 km² (38.5% of the study area), followed by the "Very High" class with 415.25 km² (38.40%), reflecting a polarized classification where extreme areas predominate. The intermediate classes, "Moderate" and "High", cover 7.03% and 7.43%, respectively, indicating little distinction in these categories. The MLP model accentuates the presence of very low susceptibility areas, reaching 478.49 km² (44.25%), but it also classifies a significant proportion of the territory as "Very high" with 449.9 km² (41.61%). The "Moderate" and "High" classes are reduced to 4% and 4.98%, suggesting a more contrasting approach, highlighting the areas with minimal and maximum risk. The XGBoost model, on the other hand, attributes 47.67% of the territory to the "Very low" class (515.4 km²), while the "Very high" class decreases to 361.58 km² (33.44%). This indicates a more balanced classification, with intermediate areas better represented: 6.64% in "Low", 5.47% in "Moderate" and 6.78% in "High".

PSO optimization modifies the distribution of classes according to the underlying model. PSO-kNN results in a more homogeneous distribution with 38.23% of the area

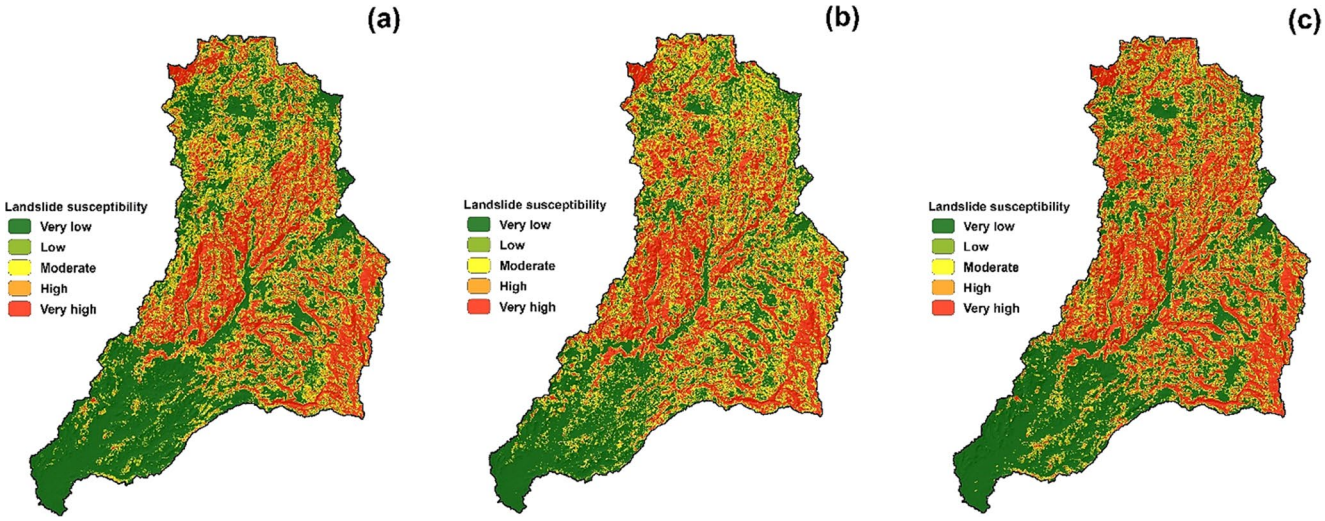

Fig. 5 Landslides susceptibility maps (**a**) XGBoost, (**b**) kNN, (**c**) MLP

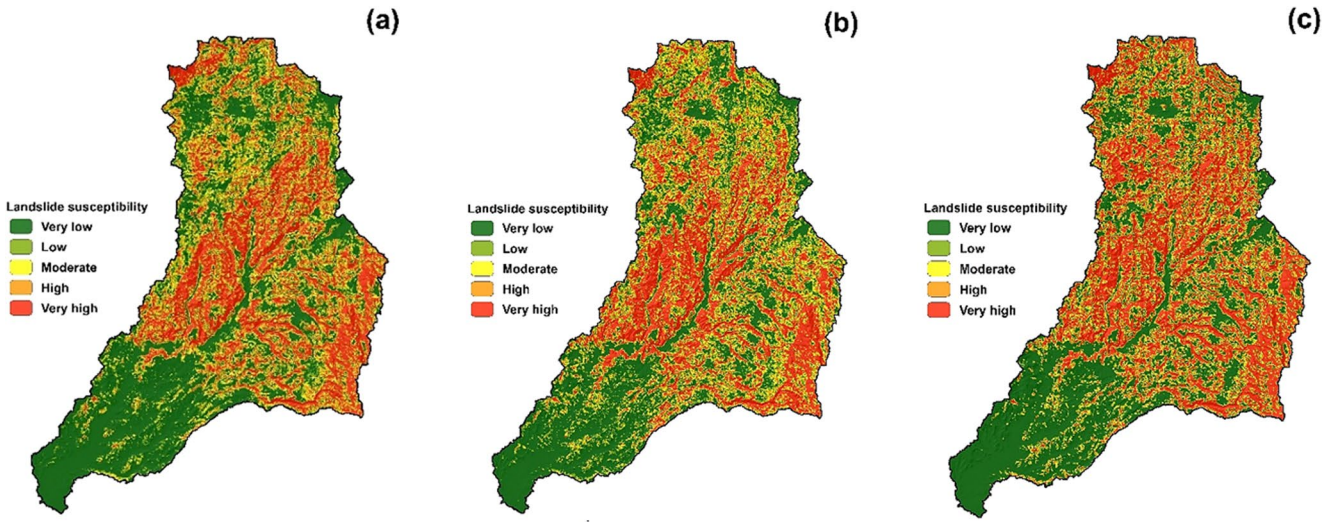

Fig. 6 Landslides susceptibility maps (**a**) PSO-XGBoost, (**b**) PSO-kNN, (**c**) PSO-ML

classified as "Very Low" (413.44 km²) and 30.31% as "Very High" (327.74 km²), while slightly increasing the proportion of intermediate classes: 10.55% as "Low", 9.99% as "Moderate" and 10.92% as "High". PSO-MLP greatly reduces the intermediate classes, with only 2.63% of the territory classified as "Moderate" and 3.13% as "High". On the other hand, it further accentuates the polarization of the extremes: 46.84% in "Very low" (506.42 km²) and 44.09% in "Very high" (476.79 km²), confirming a model that tends to favor the extreme categories to the detriment of the transition zones.

Finally, the PSO-XGBoost, while maintaining a strong presence of the "Very low" (457.87 km², 42.35%) and "Very high" (336.59 km², 31.13%) classes, offers a better distribution of the intermediate classes with 9.09% in "Low", 7.74% in "Moderate" and 9.7% in "High" (Table 4). This optimized model is distinguished by a more detailed segmentation, ensuring better representation of the moderate and high susceptibility zones.

These results demonstrate the influence of machine learning models on landslide susceptibility mapping, highlighting the contribution of optimization techniques to improve the accuracy and interpretability of classifications. PSO optimization, especially when combined with XGBoost, provides a more balanced segmentation that is faithful to real-world terrain conditions. These improvements enhance the reliability of prediction maps and their usefulness for natural hazard management, providing more nuanced and usable results for decision-makers and land use experts.

Table 4 Percentages of landslides susceptibility classes for used models

Models	Units	Very low	Low	Moderate	High	Very high
kNN	KM^2	416.35	93.38	76.01	80.36	415.25
	%	38.5	8.64	7.03	7.43	38.40
MLP	KM^2	478.49	55.86	43.21	53.81	449.9
	%	44.25	5.17	4	4.98	41.61
XGBoost	KM^2	515.4	71.840	59.11	73.31	361.58
	%	47.67	6.640	5.47	6.78	33.44
PSO-kNN	KM^2	413.44	114.03	108	118.11	327.74
	%	38.23	10.55	9.99	10.92	30.31
PSO-MLP	KM^2	506.42	35.79	28.42	33.86	476.79
	%	46.84	3.31	2.63	3.13	44.09
PSO-XGBoost	KM^2	457.87	98.27	83.7	104.83	336.59
	%	42.35	9.09	7.74	9.7	31.13

Models validation

The evaluation of the performance of different classifiers for landslide prediction highlights notable differences in terms of confusion matrix (Fig. 7), precision, recall, F1 score, accuracy, and area under the ROC curve (AUC) (Table 5). Confusion matrices (Fig. 7) show that all models performed well in predicting both positive and negative classes, with very low error rates. The PSO-XGBoost model achieved the best overall performance with the highest number of correct predictions and the lowest number of errors, followed by XGBoost and PSO-MLP. This indicates that combining PSO optimization algorithms with the baseline models contributed to improved classification accuracy compared to traditional models.

Among the tested models, the kNN algorithm exhibits a precision of 96.7%, a recall of 97.3%, and an F1 score of 97%, with an accuracy of 96.87% and an AUC of 99.1%. Although this model is effective, it remains sensitive to the data distribution, which can affect its performance.

In comparison, the MLP improves these results by achieving a precision of 97.7% and a recall of 99.2%, indicating a better generalization capability. Its F1 score reaches 98.4%, with an accuracy of 98.38% and an AUC of 99.6%, confirming its robustness for non-linear classification. However, it is the XGBoost that demonstrates the best performance among the classical models, with a precision of 99%, a recall of 99.6%, an F1 score of 99.3%, an accuracy of 99.24%, and an AUC of 99.9%. This gradient-boosting optimized algorithm excels in capturing the complex patterns present in the data, providing it with notable superiority.

The integration of optimization through PSO significantly enhances the performance of the base models. The PSO-kNN allows for a substantial improvement in results, achieving a precision of 98.8%, a recall of 99.2%, and an F1 score of 99%, while reaching an accuracy of 98.92% and an AUC of 99.4%. The impact of this optimization is even more pronounced with the PSO-MLP, which displays a precision of 98.1%, a recall of 99.2%, and an F1 score of 98.6%, with an accuracy of 98.6% and an AUC of 99.9%. These results demonstrate that PSO contributes to refining the hyperparameters of the models, thereby enhancing their classification capability.

However, the most effective model in this study remains the PSO-XGBoost, which combines the efficiency of gradient boosting and PSO optimization. With a precision of 99.2%, a recall of 99.8%, an F1 score of 99.5%, an accuracy of 99.46%, and an AUC of 99.92%, this model surpasses all other classifiers (Fig. 8). The optimization of hyperparameters via PSO reduces the risk of overfitting and maximizes the ability to distinguish between classes, making it particularly suitable for landslide prediction.

The analysis of performance improvements with PSO on the models kNN, MLP, and XGBoost highlights distinct trends in terms of impact on evaluation metrics. The PSO-kNN model shows the most significant improvements, with an increase of 2.17% in accuracy, 1.95% in recall, 2.06% in F1-score, 2.12% in precision, and 0.30% in AUC (Fig. 9). These results indicate that PSO optimization has enhanced the kNN model's ability to classify data with better accuracy and a good balance between recall and precision. On the other hand, the PSO-MLP model shows a much more modest improvement, with 0.41% in accuracy, 0.20% in F1-score, 0.22% in precision, and 0.30% in AUC. This last point means that the optimized model does not retrieve true positive classes better than the non-optimized version, which can be a problem in classification. Finally, the PSO-XGBoost model presents the weakest improvements, with increases limited to about 0.20% across all metrics, and a negligible gain in AUC (0.02%). This suggests that the XGBoost algorithm, which is already optimized by nature, benefits little from PSO optimization, unlike kNN.

PSO significantly improves the performance of kNN, but its impact is limited on MLP and XGBoost, mainly due to the intrinsic characteristics of these models. Further exploration of hyperparameters or a combination with other

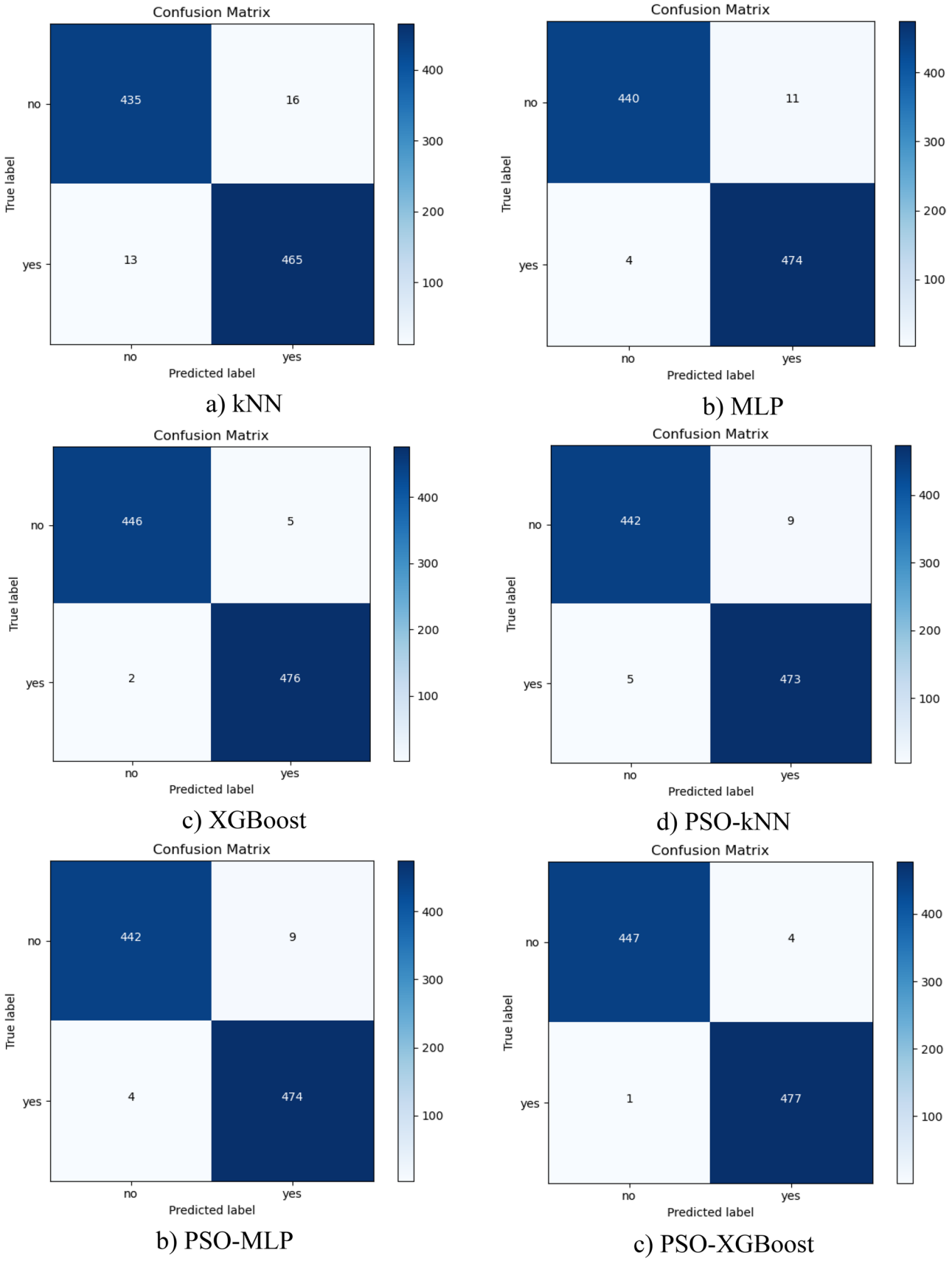

Fig. 7 Confusion matrices for comparative performance of conventional and PSO-optimized models

Table 5 Comparison of the models' outcomes

Classifier	Precision	Recall	F1 score	Accuracy	AUC
kNN	0.967	0.973	0.97	96.87	99.1
MLP	0.977	0.992	0.984	98.38	99.6
XGBoost	0.99	0.996	0.993	99.24	99.90
PSO-kNN	0.988	0.992	0.99	98.92	99.4
PSO-MLP	0.981	0.992	0.986	98.6	99.9
PSO- XGBoost	0.992	0.998	0.995	99.46	99.92

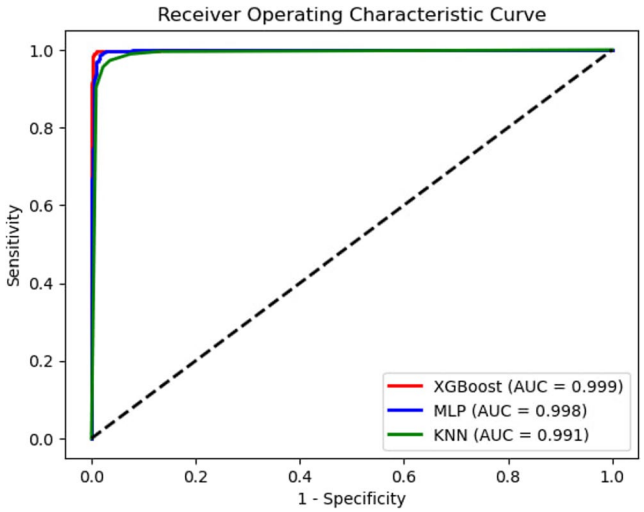

Fig. 8 ROC/AUC plot for ML models

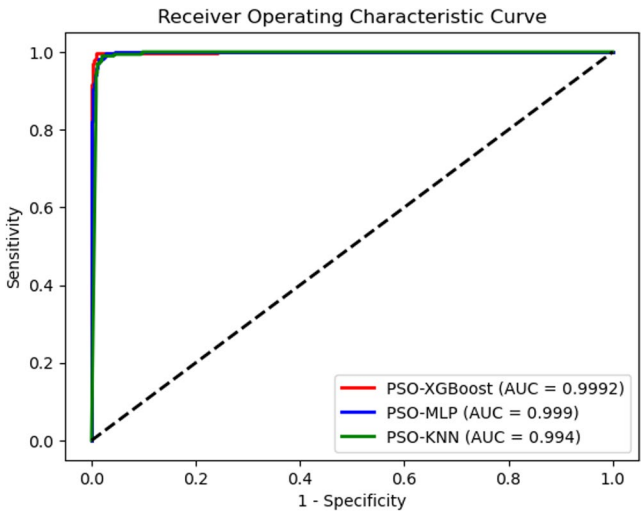

Fig. 9 ROC/AUC plot for PSO-ML models

optimization techniques could be considered to enhance the performance of MLP and XGBoost models. In conclusion, although all the models studied exhibit high performance, the results demonstrate the importance of hyperparameter optimization in improving classification. Hybrid approaches, such as PSO-XGBoost, enable achieving better precision and robustness, thereby enhancing the reliability of predictions. These results underline the relevance of advanced machine learning techniques and the contribution of optimization algorithms to refining predictive models in geospatial environments.

In order to provide further insights into the computational efficiency of the implemented models, the average training time for each algorithm was recorded. Table 6 presents the training times. As expected, the XGBoost model exhibited the longest training time due to its iterative boosting mechanism and complex hyperparameter space. The MLP model required a moderate training time, reflecting the computational cost of backpropagation and multiple hidden layers. In contrast, the kNN model demonstrated the fastest training process since no explicit model fitting is performed during training, although it is computationally more demanding during prediction. The PSO algorithm, employed for hyperparameter tuning, slightly increased the training time across all models but contributed to significant improvements in predictive performance.

In this regard, spatial uncertainty analysis provides a reliable assessment of the models used in predictions. Figure 10 illustrates the spatial distribution of uncertainty values in the predictions made by the classifiers used in this analysis. The red-shaded areas represent high uncertainty values, indicating variability in the classifier outputs in these areas. The high uncertainty values can be explained by geological complexity, topographic ruggedness, and the complex spatial interaction between other geographical factors, which affected the performance of the algorithms. Meanwhile, the green-shaded areas indicate significant consistency between the models used in this analysis, thus achieving higher reliability in predictions, with high-quality data and geological and topographical homogeneity. These results highlight the need to complement fieldwork data in areas of high uncertainty to increase the accuracy of analysis and modeling.

The ROC curve analysis reveals that the PSO-XGBoost, PSO-MLP, and PSO-kNN models exhibit remarkable performances with extremely high AUC values, illustrating their ability to effectively distinguish classes (Figs. 7 and 8). PSO-XGBoost obtains the highest AUC (0.9992, red),

Table 6 Average training time of ML models with PSO-based hyperparameter optimization for landslide susceptibility mapping

Algorithm	Training time (seconds)	Notes
kNN	32.2	Fastest training, lazy learner (costly in prediction).
MLP	72.7	Moderate time due to backpropagation and multiple hidden layers.
XGBoost	149.4	Longest due to iterative boosting and parameter optimization.
PSO	+10–15% overhead	Increased training time but improved overall performance.

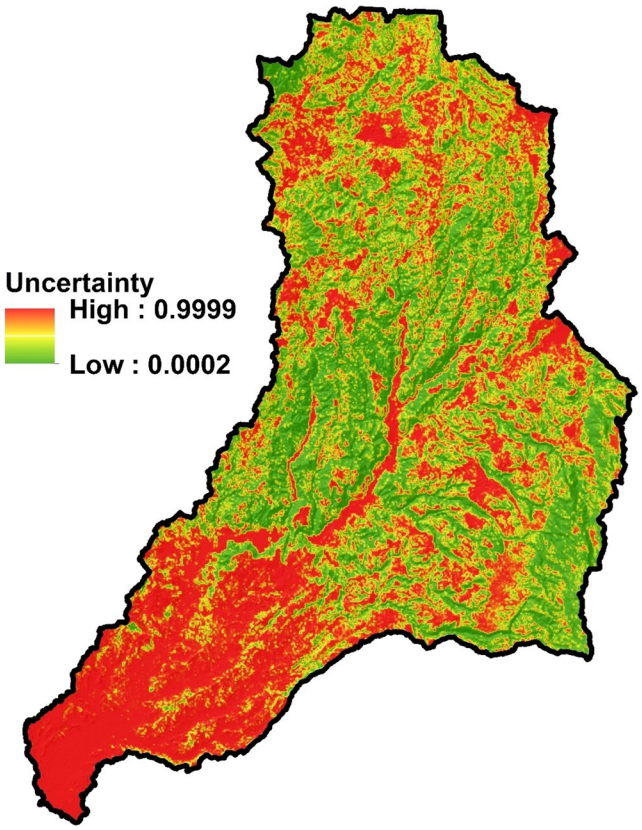

Fig. 10 Spatial variability of predictive uncertainty in mapping landslide susceptibility within the study area

demonstrating a near-perfect performance and making it the best performing model. PSO-MLP, with an AUC of 0.999 (blue), follows very closely, confirming its effectiveness after optimization. PSO-kNN, although slightly behind with an AUC of 0.994 (green), also remains a very good performing model. Overall, these results show that all three models achieve near-perfection in classification, proving the positive impact of PSO optimization. However, XGBoost stands out as the most effective model, even if the gap between the three models is minimal. This convergence of performance indicates that PSO optimization has enhanced the efficiency of each model, thus ensuring accurate and robust results for the classification studied.

Discussion

Continuously improving the performance of machine learning algorithms, whether through hybrid or optimization models, represents the cornerstone of achieving accurate predictions of landslide susceptibility. This improvement provides additional reliable dimensions for producing accurate spatial landslide risk management strategies. In the current study, we aimed to enhance the performance of kNN, MLP, and XGBoost algorithms by utilizing the PSO algorithm to improve the accuracy of landslide susceptibility prediction in the Eastern Mediterranean.

The results of this study align with previous research on landslide susceptibility, reinforcing the critical role of topographic and hydrological parameters in landslide occurrence. The Information Gain and SHAP analysis identified slope as the most dominant factor influencing landslides, a finding consistent with studies conducted in various regions worldwide. Canavesi et al. (2020); Fiolleau et al. (2023), Bhuyan et al. (2025) and Ivanik et al. (2022) also emphasized the significance of large slope ranges as the primary determinant of landslide risk. Steeper slopes increase gravitational forces, reduce soil stability, and accelerate erosion, making them highly susceptible to landslides. This universal trend highlights the necessity of including slope as a key variable in landslide susceptibility models (Cui et al. 2024; Wang et al. 2024; Wei et al. 2023).

Similarly, rainfall emerged as a crucial factor in this study, particularly due to its role in soil saturation and cohesion loss. This is consistent with findings by Abraham et al. (2020) and Yunus et al. (2021), who reported that short-duration, intense rainfall significantly contributes to landslide initiation, particularly in steep and elevated terrains. Bostan (2024) further confirmed that heavy precipitation, along with slope and proximity to faults, was among the most influential factors for landslide susceptibility in Kentucky, United States. These findings emphasize the need to incorporate rainfall variability and hydrological patterns in landslide risk assessments, especially in regions experiencing extreme weather events.

Other topographic and geological factors, such as TRI, aspect, and lithology, were also identified as significant contributors in this study. Yao et al. (2023) found that road distance, surface roughness, aspect, and land use were among the most important factors for landslide susceptibility in Lishui, Southern China. Additionally, Liu et al. 2024a, b). highlighted the significance of elevation, NDVI, distance from roads and rivers, and lithology in the Three Gorges Reservoir area of China. These findings indicate that terrain heterogeneity and geological characteristics play a fundamental role in slope stability, influencing erosion patterns, and structural weaknesses.

LULC and vegetation cover were found to have a moderate influence in this study, with NDVI showing a stabilizing effect on soil cohesion. However, its impact varied depending on environmental conditions. This is consistent with the work of Li and Tian (2025), who identified NDVI as the strongest influencing factor in their landslide susceptibility analysis. Conversely, Qiu et al. (2024) observed that higher NDVI values reduced landslide susceptibility in Shaanxi Province, China, reinforcing the protective role of

vegetation. This suggests that vegetation effects on landslide occurrence are region-specific and influenced by land management practices, climate, and soil composition.

Infrastructure and anthropogenic factors also exhibited notable effects. Distance to roads and rivers showed moderate influence in this study, which aligns with findings from Sun et al., who reported that proximity to roads, rivers, and elevation were significant predictors of landslide occurrence in the Dabashan Mountains, China. (Zheng et al. 2025) Further emphasized the impact of roads and elevations in Changbai Mountain, China, using the AutoML-SHAP approach. The presence of roads often destabilizes slopes through excavation, vibration, and drainage alterations, increasing susceptibility in developed areas.

Overall, the comparison with existing literature underscores the importance of integrating multiple environmental, topographic, hydrological, and anthropogenic factors for comprehensive landslide susceptibility assessments. While slope and rainfall consistently emerge as dominant contributors, the influence of other variables such as NDVI, land use, distance to faults, and geological features varies across regions.

To further validate the robustness of the predictive model, the AUC values obtained in this study were compared with those reported in previous research. The MLP-XGBoost model achieved an outstanding AUC of 0.9992, surpassing other machine learning and deep learning approaches used in different regions. For example, Qiu et al. (2024) obtained an AUC of 0.965 using Random Forest (RF) in Shaanxi Province, China, while Ajin et al. (2024) applied CatBoost-PSO in Kerala, India, achieving an AUC of 0.909. The CNN–DNN model employed by (Kadavi et al. 2018) in Isfahan province, Iran, produced an identical AUC of 0.909, whereas (Halder et al. 2025a, b, c) utilized Extremely Randomized Trees (ET) in West Bengal, India, with a slightly higher AUC of 0.985. The RF model used by Youssef and Pourghasemi (2021) in the Abha Basin, Saudi Arabia, demonstrated strong performance with an AUC of 0.951, while Nhu et al. (2020) achieved 0.96 using AdaBoost in Southwestern Cameron Highlands. Notably, CNN-based models such as the one by (Habumugisha et al. 2022) Habumugisha et al. (2022) in Maoxian, Sichuan Province, China, produced a lower AUC of 0.865, indicating that deep learning models may be more sensitive to regional variability in data distribution and landslide characteristics.

In addition to statistical performance metrics, the susceptibility maps were validated through intensive field surveys conducted across the study area. These surveys focused on zones previously classified as high- and very-high susceptibility in the model outputs. The field campaign involved photographic documentation, as illustrated in Fig. 11, where several mapped landslides correspond spatially with predicted high-risk zones. The outlined areas in the photographs (red polygons) delineate observed landslide scars and slope failures, providing independent empirical evidence that supports the accuracy and spatial reliability of the susceptibility maps. This integration of field-based validation strengthens

Fig. 11 Field documentation of actual landslide sites and their correspondence with high-risk zones predicted in susceptibility maps

the robustness of the modeling results and addresses the limitation of relying solely on statistical validation.

In this assessment, spatial uncertainty analysis was conducted to complement the generated landslide susceptibility maps with a measure of reliability. While susceptibility mapping highlights areas prone to landslides, it does not indicate the degree of confidence in the predictions. The uncertainty assessment was therefore necessary to identify zones of agreement and disagreement among the applied classifiers. Areas of low uncertainty, where the models converged towards similar outputs, were interpreted as reliable predictions supported by consistent geological and topographical conditions. Conversely, areas of high uncertainty reflected divergence among the models, which can be attributed to geological complexity, rugged terrain, and the interaction of multiple environmental factors. These zones mark priority areas where complementary field surveys or additional high-resolution data are required to improve model accuracy. By integrating uncertainty analysis, the study provides not only susceptibility predictions but also practical guidance on the reliability of the outputs, thereby supporting more informed decision-making for landslide risk management.

The superior performance of the MLP-PSO model in this study suggests that hybrid optimization techniques effectively enhance the predictive capabilities of machine learning models by optimizing hyperparameters and improving feature selection. The integration of multiple environmental variables, including topographic, hydrological, geological, and anthropogenic factors, has contributed to the high accuracy of the model. These findings emphasize the importance of using advanced ML techniques alongside explainability tools like SHAP to better understand the relative contributions of different factors to landslide susceptibility.

Limitation

Despite the strong predictive performance and comprehensive validation framework employed in this study, several limitations should be acknowledged. First, the quality and resolution of the input datasets, particularly rainfall, lithology, and LULC data, may introduce uncertainty into the modeling process. Spatial and temporal inconsistencies in these datasets can affect the representation of triggering mechanisms, especially in complex mountainous terrains. Second, although PSO optimization significantly improved model performance, the computational cost associated with hybrid models may limit their applicability for large-scale or real-time assessments. Additionally, the models were trained using historical landslide inventories, which may contain positional inaccuracies, omission errors, or temporal biases, potentially influencing prediction reliability.

Moreover, the transferability of the developed models to other regions remains limited, as landslide susceptibility is highly dependent on local geological, climatic, and geomorphological conditions. The relative importance of conditioning factors identified in this study may therefore vary in different environmental settings. Although uncertainty analysis and field validation were conducted, some areas with high uncertainty persist, particularly in zones characterized by complex geology and intense human activity. Finally, this study focused on susceptibility mapping rather than dynamic landslide hazard assessment, and thus does not explicitly account for temporal triggering processes or future climate change scenarios. Addressing these limitations in future research through higher-resolution datasets, multi-temporal modeling, and climate-informed scenarios would further enhance the robustness and applicability of landslide susceptibility assessments.

Conclusion and future recommendations

This study demonstrates the strong potential of machine learning–based approaches, particularly the PSO-optimized XGBoost model, for landslide susceptibility mapping in complex mountainous environments. The results confirm that topographic and hydrological factors—most notably slope, TRI, rainfall, aspect, and TWI—play a dominant role in controlling landslide occurrence. The integration of the PSO algorithm effectively improved model performance through hyperparameter optimization, with PSO-XGBoost achieving the highest performance. However, these results should be interpreted in light of data resolution, inventory uncertainty, and the region-specific nature of landslide processes, which may influence model generalization beyond the study area.

Feature importance analysis using SHAP identified slope as the most influential factor, followed by moderate contributions from lithology, NDVI, and LULC. While proximity to roads and rivers, flow accumulation, and soil depth exhibited comparatively lower importance, their influence remains condition-dependent, particularly in areas affected by intense anthropogenic activity or localized hydrological concentration. This highlights the necessity of integrating multiple conditioning factors to adequately capture the complexity of landslide mechanisms.

Future research should prioritize the use of higher-resolution and multi-source datasets, including LiDAR and high-resolution satellite imagery, to better characterize terrain heterogeneity and reduce spatial uncertainty. Incorporating long-term climatic trends and rainfall variability, as well as multi-temporal and time-series modeling approaches, would enhance the representation of dynamic triggering processes. In addition, hybrid artificial intelligence

frameworks that combine deep learning architectures with traditional machine learning models could improve feature extraction and model transferability. Continued field-based validation remains essential to strengthen model reliability and address uncertainties in complex geological settings. Finally, embedding susceptibility models within GIS-based decision-support systems would improve their practical applicability for land-use planning, disaster risk reduction, and early warning strategies.

Overall, while PSO-optimized XGBoost proved to be a robust and high-performing approach for landslide susceptibility assessment, further efforts are required to enhance its scalability, temporal adaptability, and applicability under changing climatic and environmental conditions.

Acknowledgements The authors extend their appreciation to the Deanship of Scientific Research at King Khalid University for funding this work through Research Group under grant number RGP2/599/46.

Author contributions Hazem Ghassan Abdo: Data curation, Investigation, Visualization, Writing – original draft. Sahar Mohammed Richi: Data curation, Investigation, Visualization. Bilel Zerouali: Writing – review & editing. Okan Mert katipoğlu: Methodology, Writing – review & editing. Pankaj Prasad: Writing – review & editing. Hasan Arman: Writing – review & editing. Javed Mallick and Jasem A Albanai: Data curation. All authors read and approved the final manuscript.

Data availability The data that support the findings of this study are available on request from the corresponding author.

Declarations

Ethics approval Not applicable.

Financial interests The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Consent to participate Not applicable.

Consent to publish Not applicable.

Competing interests The authors declare no competing interests.

References

Abraham MT, Satyam N, Rosi A, Pradhan B, Segoni S (2020) The selection of rain gauges and rainfall parameters in estimating intensity-duration thresholds for landslide occurrence: case study from wayanad (India). Water 12(4):1000

Ahmad RA, Singh RP, Adris A (2017) Seismic hazard assessment of Syria using seismicity, DEM, slope, active faults and GIS. Remote Sens Applications: Soc Environ 6:59–70

Ajin RS, Segoni S, Fanti R (2024) Optimization of SVR and CatBoost models using metaheuristic algorithms to assess landslide susceptibility. Sci Rep 14(1):24851

Al-Najjar HAH, Pradhan B, Beydoun G, Sarkar R, Park HJ, Alamri A (2023) A novel method using explainable artificial intelligence (XAI)-based shapley additive explanations for spatial landslide prediction using time-series SAR dataset. Gondwana Res 123:107–124. https://doi.org/10.1016/j.gr.2022.08.004

Al-Rawabdeh A, Awawdeh M, Al Quraan H, Jaradat R (2024) Modeling landslides hazard along Amman–Jerash–Irbid Highway, Jordan by integrating open street map (OSM) and weighted linear combination (WLC) techniques. Modeling Earth Systems and Environment 10(2):2547–2565

Al-Shabeeb AR, Al-Fugara AK, Khedher KM, Mabdeh AN, Al-Adamat R (2022) Spatial mapping of landslide susceptibility in Jerash Governorate of Jordan using genetic algorithm-based wrapper feature selection and bagging-based ensemble model. Geomat Nat Hazards Risk 13(1):2252–2282

Alemayehu D, Tadesse M, As MAA (2018) Application of TPI for analysis of landforms and LULC of Adama Wereda, Ethiopia. J Geograph Stud 2:100–109

Ali SA, Parvin F, Vojtekova J, Costache R, Linh NTT, Pham QB, Vojtek M, Gigovic L, Ahmad A, Ghorbani MA (2021) GIS-based landslide susceptibility modeling: a comparison between fuzzy multi-criteria and machine learning algorithms. Geosci Front 12:857–876

Aslam B, Maqsoom A, Khalil U, Ghorbanzadeh O, Blaschke T, Farooq D, Tufail RF, Suhail SA, Ghamisi P (2022) Evaluation of different landslide susceptibility models for a local scale in the Chitral District. Northern Pakistan. Sensors 22(9):3107. https://doi.org/10.3390/s22093107

Awawdeh MM, ElMughrabi MA, Atallah MY (2018) Landslide susceptibility mapping using GIS and weighted overlay method: a case study from North Jordan. Environ Earth Sci 77:1–15

Badreldin H, Scaini C, Hassan HM, Peresan A (2025) High-resolution multi-hazard residential buildings and population exposure model for coastal areas: a case study in northeastern Italy. Int J Disaster Risk Reduct 121:105403

Bansal M, Goyal A, Choudhary A (2022) A comparative analysis of K-nearest neighbor, genetic, support vector machine, decision tree, and long short term memory algorithms in machine learning. Decis Anal J 3:100071

Bhuyan K, Rana K, Ozturk U, Nava L, Rosi A, Meena SR, Fan X, Floris M, van Westen C, Catani F (2025) Towards automatic delineation of landslide source and runout. Eng Geol 345:107866

Bostan T (2024) Generating a landslide susceptibility map using integrated meta-heuristic optimization and machine learning models. Sustainability 16(21):9396

Bravo-López E, Fernández Del Castillo T, Sellers C, Delgado-García J (2023) Analysis of conditioning factors in cuenca, Ecuador, for landslide susceptibility maps generation employing machine learning methods. Land 12(6):1135

Canavesi V, Segoni S, Rosi A, Ting X, Nery T, Catani F, Casagli N (2020) Different approaches to use morphometric attributes in landslide susceptibility mapping based on meso-scale

spatial units: a case study in Rio de Janeiro (Brazil). Remote Sens 12(11):1826
Chen T, Guestrin C (2016) XGBoost: A scalable tree boosting system. https://doi.org/10.1145/2939672.2939785
Chen W, Chen X, Peng J, Panahi M, Lee S (2021) Landslide susceptibility modeling based on ANFIS with teaching-learning-based optimization and satin bowerbird optimizer. Geosci Front 12:93–107
Cui HZ, Tong B, Wang T, Dou J, Ji J (2024) A hybrid data-driven approach for rainfall-induced landslide susceptibility mapping: physically-based probabilistic model with convolutional neural network. J Rock Mech Geotech Eng 17(8):4933–4951
Dai X, Chen J, Zhang T, Xue C (2025) Integrated landslide risk assessment via a landslide susceptibility model based on intelligent optimization algorithms. Remote Sens 17(3):545
Dey S, Das S (2025) Swarm optimization based heterogeneous machine learning techniques for enhanced landslide susceptibility assessment with comprehensive uncertainty quantification. Earth Sci Inf 18(1):145
Fang Z, Wang Y, Peng L, Hong H (2020) Integration of convolutional neural network and conventional machine learning classifiers for landslide susceptibility mapping. Comput Geosci 139:104470. https://doi.org/10.1016/j.cageo.2020.104470
Fiolleau S, Uhlemann S, Falco N, Dafflon B (2023) Assessing probability of failure of urban landslides through rapid characterization of soil properties and vegetation distribution. Geomorphology 423:108560
Francone L (2022) Vulnerability assessment of buildings to landslides (Doctoral dissertation, Politecnico di Torino)
Froude MJ, Petley DN (2018) Global fatal landslide occurrence from 2004 to 2016. Nat Hazards Earth Syst Sci 18(8):2161–2181. https://doi.org/10.5194/nhess-18-2161-2018
Guzzetti F, Carrara A, Cardinali M, Reichenbach P (1999) Landslide hazard evaluation: a review of current techniques and their application in a multi-scale study, central Italy. Geomorphology 31:181–216
Guzzetti F, Mondini AC, Cardinali M, Fiorucci F, Santangelo M, Chang KT (2012) Landslide inventory maps: new tools for an old problem. Earth-Sci Rev 112:42–66
Habumugisha JM, Chen N, Rahman M, Islam MM, Ahmad H, Elbeltagi A, Sharma G, Liza SN, Dewan A (2022) Landslide susceptibility mapping with deep learning algorithms. Sustainability 14(3):1734
Halder K, Srivastava AK, Ghosh A, Das S, Banerjee S, Pal SC, Chatterjee U, Bisai D, Ewert F, Gaiser T (2025a) Improving landslide susceptibility prediction through ensemble recursive feature elimination and meta-learning framework. Sci Rep 15(1):5170
Halder K, Srivastava AK, Ghosh A, Das S, Banerjee S, Pal SC, Gaiser T (2025b) Improving landslide susceptibility prediction through ensemble recursive feature elimination and meta-learning framework. Sci Rep 15(1):5170
Halder K, Srivastava AK, Ghosh A, Das S, Banerjee S, Pal SC, Chatterjee U, Bisai D, Ewert F, Gaiser T (2025c) Improving landslide susceptibility prediction through ensemble recursive feature elimination and meta-learning framework. Sci Rep 15(1):5170. https://doi.org/10.1038/s41598-025-87587-3
Hallal N, Hamidatou M, Medjnoun A, Hamai L, Lamali A, Hassan HM, Fahem D (2024) GIS-based statistical and limit equilibrium models in the assessment of slope stability and landslide susceptibility: the case study of the Aomar Miocene basin, Bouira, Algeria. Environ Earth Sci 83(20):578
Hong HY, Pourghasemi HR, Pourtaghi ZS (2016) Landslide susceptibility assessment in Lianhua County (China): a comparison between a random forest data mining technique and bivariate and multivariate statistical models. Geomorphology 259:105–118
Hong HY, Shahabi H, Shirzadi A, Chen W, Chapi K, Ahmad B, Roodposhti B, Hesar MS, Tian AY, Bui YY (2019b) Landslide susceptibility assessment at the Wuning area, China: a comparison between multi-criteria decision making, bivariate statistical and machine learning methods. Nat Hazards 96:173–212
Hou M, Wang Y, Bai X, Yuan R (2025) Evaluation of landslide susceptibility in the northern section of the Xiaojiang fault zone based on factor optimization. Landslides. https://doi.org/10.1007/s10346-024-02448-w
Huang S, Tian N, Wang Y, Ji Z (2016) Particle swarm optimization using multi-information characteristics of all personal-best information. SpringerPlus 5:1632
Iheanetu K, Obileke K (2024) Short-term forecasting of photovoltaic power using multilayer perceptron neural network, convolutional neural network, and k-nearest neighbors' algorithms. Opt 5(2):293–309
Ivanik O, Fonseca J, Shabatura O, Khomenko R, Hadiatska K, Kravchenko D (2022) An integrated approach for landslide hazard assessment: a case study of the middle Dnieper Basin, Ukraine. J Water Land Dev 52:81–86
Ji S, Yu D, Shen C, Li W, Xu Q (2020) Landslide detection from an open satellite imagery and digital elevation model dataset using attention boosted convolutional neural networks. Landslides 17:1337–1352. https://doi.org/10.1007/s10346-020-01353-2
Jiang W, Li L, Niu R (2025) Impact of non-landslide sample sampling strategies and model selection on landslide susceptibility mapping. Appl Sci 15(4):2132
Kadavi PR, Lee C-W, Lee S (2018) Application of ensemble-based machine learning models to landslide susceptibility mapping. Remote Sens 10(8):1252
Kalantar B, Ueda N, Saeidi V, Ahmadi K, Halin AA, Shabani F (2020) Landslide susceptibility mapping: machine and ensemble learning based on remote sensing big data. Remote Sens 12(11):1737. https://doi.org/10.3390/rs12111737
Karakas G, Unal EO, Cetinkaya S, Ozcan NT, Karakas VE, Can R, Gokceoglu C, Kocaman S (2024) Analysis of landslide susceptibility prediction accuracy with an event-based inventory: the 6 February 2023 Turkiye earthquakes. Soil Dyn Earthq Eng 178:108491
Li M, Tian H (2025) Insights from optimized non-landslide sampling and SHAP explainability for landslide susceptibility prediction. Appl Sci 15(3):1163
Liu B, Guo H, Li J, Ke X, He X (2024a) Application and interpretability of ensemble learning for landslide susceptibility mapping along the three Gorges reservoir area, China. Nat Hazards 120(5):4601–4632
Liu S, Wang L, Zhang W, Sun W, Wang Y, Liu J (2024b) Physics-informed optimization for a data-driven approach in landslide susceptibility evaluation. J Rock Mech Geotech Eng 16(8):3192–3205
Lu H, Chen J, Guo L (2018) Energy quality management. Comprehensive Energy Systems. pp 258–314. https://doi.org/10.1016/B978-0-12-809597-3.00521-6
Lv L, Chen T, Dou J, Plaza A (2022) A hybrid ensemble-based deep-learning framework for landslide susceptibility mapping. Int J Appl Earth Obs Geoinf 108:102713. https://doi.org/10.1016/j.jag.2022.102713
Meng S, Shi Z, Li G, Peng M, Liu L, Zheng H, Zhou C (2024) A novel deep learning framework for landslide susceptibility assessment using improved deep belief networks with the intelligent optimization algorithm. Comput Geotech 167:106106
Meshram SG, Meshram C, Pourhosseini FA, Hasan MA, Islam S (2022) A multi-layer perceptron (MLP)-Fire fly algorithm (FFA)-based model for sediment prediction. Soft Computing, pp 1–10

Moore ID, Wilson JP (1992) Length-slope factors for the revised universal soil loss equation - simplified method of estimation. J Soil Water Conserv 47:423–428

Moosavi V, Niazi Y (2016) Development of hybrid wavelet packet-statistical models (WP-SM) for landslide susceptibility mapping. Landslides 13:97–114

Nhu V-H, Mohammadi A, Shahabi H, Bin B, Al-Ansari N, Shirzadi A, Clague JJ, Jaafari A, Chen W, Nguyen H (2020) Landslide susceptibility mapping using machine learning algorithms and remote sensing data in a tropical environment. Int J Environ Res Public Health 17(14):4933

Pawluszek-Filipiak K, Borkowski A (2020) On the importance of train–test split ratio of datasets in automatic landslide detection by supervised classification. Remote Sens 12(18):3054

Pham BT, Bui DT, Dholakia MB, Prakash I, Pham HV, Mehmood K, Le HQ (2017) A novel ensemble classifier of rotation forest and Naïve bayer for landslide susceptibility assessment at the Luc Yen district, Yen Bai Province (Viet Nam) using GIS. Geomat Nat Hazards Risk 8(2):649–671. https://doi.org/10.1080/19475705.2016.1255667

Pradhan AMS, Kim YT (2014) Relative effect method of landslide susceptibility zonation in weathered granite soil: a case study in Deokjeok-ri creek, South Korea. Nat Hazards 72:1189–1217

Prasad P, Loveson VJ, Das B, Kotha M (2022) Novel ensemble machine learning models in flood susceptibility mapping. Geocarto International, pp 1–22. https://doi.org/10.1080/10106049.2021.1892209

Qiu H, Xu Y, Tang B, Su L, Li Y, Yang D, Ullah M (2024) Interpretable landslide susceptibility evaluation based on model optimization. Land 13(5):639

Rabby YW, Li Y, Abedin J, Sabrina S (2022) Impact of land use/land cover change on landslide susceptibility in Rangamati municipality of Rangamati district, Bangladesh. ISPRS International journal of geo-information 11(2):89

Riley SJ, Degloria SD (1999) A Terrain ruggedness index that quantifies topographic heterogeneity in pursuit of wild game view project nonpoint source pollution in China view project. https://www.researchgate.net/publication/259011943

Sahab MG, Toropov VV, Gandomi AHA (2013) Review on traditional and modern structural optimization. Metaheuristic Appl Struct Infrastruct 25–47. https://doi.org/10.1016/B978-0-12-398364-0.00002-4

Sahana M, Pham BT, Shukla M, Costache R, Thu DX, Chakrabortty R, Prakash I (2022) Rainfall induced landslide susceptibility mapping using novel hybrid soft computing methods based on multi-layer perceptron neural network classifier. Geocarto Int 37(10):2747–2771

Segoni S, Ajin RS, Nocentini N, Fanti R (2024) Insights gained from the review of landslide susceptibility assessment studies in Italy. Remote Sens 16:1–31

Selamat SN, Majid NA, Taha MR (2025) Multicollinearity and spatial correlation analysis of landslide conditioning factors in langat river Basin, Selangor. Nat Hazards 121(3):2665–2684

Shirzadi A, Soliamani K, Habibnejhad M, Kavian A, Chapi K, Shahabi H, Chen W, Khosravi K, Pham T, Pradhan B, Ahmad A, Ahmad B, Bui T (2018) Novel GIS based machine learning algorithms for shallow landslide susceptibility mapping. Sensors 18(11):3777–3805

Singh A, Dhiman N, Kc N, Shukla DP (2024) Improving ML-based landslide susceptibility using ensemble method for sample selection: a case study of Kangra district in Himachal Pradesh, India. Environ Sci Pollut Res Int. https://doi.org/10.1007/s11356-024-34726-4

Sun DL, Wen HJ, Wang DZ, Xu JH (2020b) A random forest model of landslide susceptibility mapping based on hyperparameter optimization using Bayes algorithm. Geomorphology 362:107201

Sun DL, Xu JH, Wen HJ, Wang Y (2020c) An optimized random forest model and its generalization ability in landslide susceptibility mapping: application in two areas of Three Gorges Reservoir. China. J. Earth Sci-China 31:1068–1086

Sun DL, Xu JH, Wen HJ, Wang DZ (2021) Assessment of landslide susceptibility mapping based on Bayesian hyperparameter optimization: a comparison between logistic regression and random forest. Eng Geol 281:105972

Susena Y, Hadmoko DS, Wibowo SB (2025) Machine learning techniques on spatio-temporal data for landslide susceptibility assessment at Dieng Mountainous Region, Banjarnegara district, Central Java, Indonesia. Nat Hazards. https://doi.org/10.1007/s11069-025-07136-z

Taher M, Mourabit T, Talibi E, Amine H, Bourjila A, Errahmouni A, Etebaai A, I (2025) Landslide susceptibility mapping (LSM) of the boudinar basin (Morocco) using the geographic information system (GIS) and the analytical hierarchy process (AHP) method. Iran J Earth Sci 17(1):1–10

Tanoumand A, Mashayekhi M, Majdi A, S Ramadan M (2025) Development of a method for optimal location of emergency water reservoirs for FFE suppression using AHP and GIS. Front Built Environ 11:1638961

Tebbouche MY, Benamar A, Hassan D, Singh HM, Bencharif AP, Machane R, Nemer D, Z (2022) Characterization of El Kherba landslide triggered by the August 07, 2020, M w = 4.9 Mila earthquake (Algeria) based on post-event field observations and ambient noise analysis. Environ Earth Sci 81(2):46

Tekin S, Roman AQ, Can T (2024) Landslide susceptibility assessment of the Asi watershed, Southern Türkiye. Turk J Earth Sci 33(2):208–223

Tyagi A, Tiwari RK, James N (2023) Mapping the landslide susceptibility considering future land-use land-cover scenario. Landslides 20(1):65–76

Wang Y, Fang ZC, Hong HY (2019) Comparison of convolutional neural networks for landslide susceptibility mapping in Yanshan County, China. Sci Total Environ 666:975–993

Wang YH, Wang LQ, Zhang WG, Liu SL, Sun WX, Hong L, Zhu ZW (2024) A physics-informed machine learning solution for landslide susceptibility mapping based on three-dimensional slope stability evaluation. J Cent South Univ 31(11):3838–3853

Wei X, Zhang L, Luo J, Liu D (2021) A hybrid framework integrating physical model and convolutional neural network for regional landslide susceptibility mapping. Nat Hazards 109(1):471–497

Wei X, Zhang L, Gardoni P, Chen Y, Tan L, Liu D, Li H (2023) Comparison of hybrid data-driven and physical models for landslide susceptibility mapping at regional scales. Acta Geotech 18(8):4453–4476

Wei X, Gardoni P, Zhang L, Tan L, Liu D, Du C, Li H (2024) Improving pixel-based regional landslide susceptibility mapping. Geosci Front 15(4):101782

Wu Z, Ye R, Yang S, Wen T, Huang J, Chen Y (2024) Study on early identification of rainfall-induced accumulation landslide hazards in the Three Gorges Reservoir Area. Remote Sens 16(10):1669

Yang Y, Peng S, Huang B, Xu D, Yin Y, Li T, Zhang R (2024) Multiscale analysis of the susceptibility of different landslide types and identification of the main controlling factors. Ecol Indic 168:112797

Yao Z, Chen M, Zhan J, Zhuang J, Sun Y, Yu Q, Yu Z (2023) Refined landslide susceptibility mapping by integrating the shap-catboost model and insar observations: a case study of Lishui, Southern China. Appl Sci 13(23):12817

Yi Y, Zhang Z, Zhang W, Jia H, Zhang J (2020) Landslide susceptibility mapping using multiscale sampling strategy and convolutional neural network: a case study in Jiuzhaigou region. Catena 195:104851

Youssef AM, Pourghasemi HR (2021) Landslide susceptibility mapping using machine learning algorithms and comparison of their performance at Abha Basin, Asir Region, Saudi Arabia. Geosci Front 12(2):639–655

Yu H, Pei W, Zhang J, Chen G (2023) Landslide susceptibility mapping and driving mechanisms in a vulnerable region based on multiple machine learning models. Remote Sens 15(7):1886

Yuan X, Lu H, Cheng J, Nie R, Xu Y, Zhang J, Ma L, Liu C, Dai X, Fu X, Li W, Tang M, Yang Z (2022) A comparative analysis of certainty factor-based machine learning methods for collapse and landslide susceptibility mapping in Wenchuan County, China. Remote Sens 14(14):3259. https://doi.org/10.3390/rs14143259

Yunus AP, Fan X, Subramanian SS, Jie D, Xu Q (2021) Unraveling the drivers of intensified landslide regimes in Western Ghats, India. Sci Total Environ 770:145357

Zeng T, Guo Z, Wang L, Jin B, Wu F, Guo R (2023a) Tempo-spatial landslide susceptibility assessment from the perspective of human engineering activity. Remote Sens 15(16):4111

Zeng T, Wu L, Peduto D, Glade T, Hayakawa YS, Yin K (2023b) Ensemble learning framework for landslide susceptibility mapping: different basic classifier and ensemble strategy. Geosci Front 14(6):101645

Zhang SH, Li ZH, Hou XN, Yi YJ (2019) Impacts on watershed-scale runoff and sediment yield resulting from synergetic changes in climate and vegetation. CATENA 179:129–138

Zheng D, Li Y, Yan C, Wu H, Yamashiki YA, Gao B, Nian T (2025) Landslide susceptibility assessment using AutoML-SHAP method in the southern foothills of Changbai Mountain, China. Landslides. https://doi.org/10.1007/s10346-025-02462-6

Zhou X, Wen H, Zhang Y, Xu J, Zhang W (2021) Landslide susceptibility mapping using hybrid random forest with GeoDetector and RFE for factor optimization. Geosci Front 12(5):101211

Authors and Affiliations

Hazem Ghassan Abdo[1,2] · Sahar Mohammed Richi[3] · Bilel Zerouali[4] · Saeed Alqadhi[5] · Okan Mert Katipoğlu[6] · Pankaj Prasad[7] · Hasan Arman[8] · Jasem A Albanai[9] · Javed Mallick[5]

✉ Hazem Ghassan Abdo
habdo@ogs.it

Sahar Mohammed Richi
sahar1.richi@damascusuniversity.edu.sy

Bilel Zerouali
b.zerouali@univ-chlef.dz

Saeed Alqadhi
sdalqadi@kku.edu.sa

Okan Mert Katipoğlu
okatipoglu@erzincan.edu.tr

Pankaj Prasad
ppankaj.earthscience@gmail.com

Hasan Arman
harman@uaeu.ac.ae

Jasem A Albanai
jasem.albanai@ouce.ox.ac.uk

Javed Mallick
jmallick@kku.edu.sa

[1] National Institute of Oceanography and Applied Geophysics – OGS, Borgo Grotta Gigante 42/C, Sgonico (TS) 34010, Italy

[2] Geography Department, Faculty of Arts and Humanities, Tartous University, Tartous, Syria

[3] Geography Department, Faculty of Arts and Humanities, Damascus University, Damascus, Syria

[4] Laboratory of Architecture, Cities and Environment, Department of Hydraulic, Faculty of Civil Engineering and Architecture, Hassiba Benbouali University of Chlef, B.P. 78C, Ouled Fares, Chlef 02180, Algeria

[5] Department of Civil Engineering, College of Engineering, King Khalid University, Abha, Kingdom of Saudi Arabia

[6] Department of Civil Engineering, Erzincan Binali Yıldırım University, Erzincan 24002, Turkey

[7] Geoinformatics division, Department of Geography, Mahapurusha Srimanta Sankaradeva University, Nagaon, Assam 782001, India

[8] Geosciences Department, National Water and Energy Center (NWEC), United Arab Emirates University, Al Ain 15551, United Arab Emirates

[9] School of Geography and the Environment, University of Oxford, Oxford OX1 2JD, UK