BILINGUAL PUBLISHING GROUP

Journal of Environmental & Earth Sciences

https://journals.bilpubgroup.com/index.php/jees

ARTICLE

Ensemble Machine Learning Applied to Assessment and Mapping of Low and Moderate Slopes Landslide Susceptibility in Hamam Nbail, Algeria

Harfouche Ameur [*] **, *Djerbal Lynda*** **, *Bahar Ramdane***

Department of Civil Engineering, Laboratory of LEEGO, University of Sciences and Technology Houari Boumediene (USTHB), Algiers 16111, Algeria

ABSTRACT

The municipality of Hammam N'bails, located 37 km east of the capital of Guelma province (eastern Algeria), is accessible via RN20 and CW19 roads. It borders the municipalities of Khemissa and El Henancha in Souk-Ahras province. With a population of approximately 16,000 and covering an area of 164 km², this region is characterized by mountainous terrain, with elevations ranging from 112 to 292 meters. The area experiences cold, snowy winters and hot, dry summers, with an average annual rainfall of about 600 mm. Renowned for its natural thermal springs, Hammam N'bails is also a notable tourist destination. The rugged topography of the region leads to frequent landslides, particularly on medium and low slopes. Landslide susceptibility is assessed using raster calculators in ArcGIS and efficient machine learning algorithms, such as Decision Trees, Bagging, Random Forest, SVM, and MLP. Factors considered in the analysis include slope, elevation, geology, aspect, proximity to streams and roads, land cover, and rainfall. The performance of these models is evaluated using ROC-AUC curves, providing a robust method to understand and mitigate geological risks in this area.

Keywords: Landslide; Low and Middle Slopes; Susceptibility; Machine Learning; Hamam Nbail

*CORRESPONDING AUTHOR:

Harfouche Ameur, Department of Civil Engineering, Laboratory of LEEGO, University of Sciences and Technology Houari Boumediene (USTHB), Algiers 16111, Algeria; Email: ameur.harfouche@usthb.dz

ARTICLE INFO

Received: 6 November 2024 | Revised: 29 November 2024 | Accepted: 4 December 2024 | Published Online: 7 February 2025
DOI: https://doi.org/10.30564/jees.v7i2.7699

CITATION

Ameur, H., Lynda, D., Ramdane, B., 2025. Ensemble Machine Learning Applied to Assessment and Mapping of Low and Moderate Slopes Landslide Susceptibility in Hamam Nbail, Algeria. Journal of Environmental & Earth Sciences. 7(2): 161–174. DOI: https://doi.org/10.30564/jees.v7i2.7699

1. Introduction

Landslides pose a significant and complex global threat, endangering lives, infrastructure, and transportation networks[1–3]. Each year, slope failures cause substantial direct and indirect economic damages, often amounting to millions of dollars[4, 5]. In the past two decades, the development of global landslide susceptibility maps has become an invaluable tool for identifying landslide-prone areas in both urban and rural settings

Numerous methodologies have been developed for evaluating and mapping landslide susceptibility[6, 7], encompassing probabilistic models such as frequency ratio, weight of evidence[8]; logistic regression model[9]; Didier and Tritch model[10], Djerbal and Melbouci model[1], and data mining utilizing fuzzy logic and artificial neural networks[11].

Qualitative methods are based on direct field observations and expert judgment rooted in experience, providing straightforward techniques for assessing landslide susceptibility. In contrast, quantitative methods use statistical and mathematical approaches, offering objective ways to analyze landslides[12].

Similarly, quantitative methods can be classified into several categories, including statistical, deterministic, and machine learning approaches[12]. Deterministic methods, also known as geotechnical methods, have been extensively utilized in research. These methods rely on site-specific geotechnical parameters and engineering principles of slope instability, often expressed in terms of safety factors. However, deterministic methods have limitations as they require detailed geotechnical and hydrological data, which can be challenging to collect over large areas, and they are typically suitable for mapping small regions[13].

The emergence of machine learning has revolutionized landslide prediction by leveraging algorithms to analyze large datasets and identify complex patterns contributing to slope instability. Machine learning models, such as decision trees and neural networks, enhance accuracy by integrating various environmental and terrain variables, thereby improving the understanding and prediction of landslide occurrences with higher precision and efficiency.

Moderate and low slopes can experience significant landslides covering extensive areas. Therefore, modeling these terrains involves either observing the high degree of low slope across the entire area with the potential for landslides on these slopes or studying the overall susceptibility and utilizing GIS tools to assess specific susceptibility in these areas.

The objective of this approach is to demonstrate that moderate slope terrains, while not showing specific landslides, can still be susceptible based on spatial analysis and numerical modeling provided by machine learning.

Determining landslide susceptibility in moderate and low slopes is crucial for public decision-makers. It allows for planning public infrastructure that cannot, due to safety and accessibility reasons, be feasibly constructed on steep terrain. This approach ensures safer and more accessible development projects in suitable areas.

2. Materials and Methods

2.1. Case Study

The municipality of Hammam N'bails (**Figure 1**) is located to the east of the capital of the Guelma province (**Figure 2**), 37 km away from Guelma-ville via RN20 and then CW19. It shares borders with the municipalities of Khemissa and El Henancha in the Souk-Ahras province.

Figure 1. Hamam Nbail.

Hammam N'bails, with a population of approximately 16,000 inhabitants and covering an area of 164 km², is characterized by mountainous terrain with elevations ranging from 112 to 292 meters. The region experiences cold and snowy winters and hot, dry summers, with an average annual rainfall of around 600 mm. Known for its natural thermal springs, the area is also recognized for its tourism activities.

The topography, characterized by steep slopes reaching up to 62 degrees, and the soil cover consisting of recent formations from the continental Cretaceous and Pontian pe-

riods, confirm a high susceptibility to landslides. These have been identified both on-site and through the consultation of satellite photos from Google Earth.

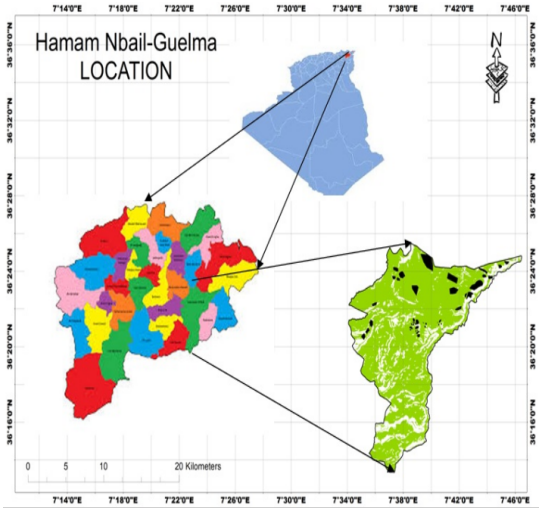

Figure 2. Location map.

2.2. Thematic Map Layers

Initially, Google Earth imagery was used to identify areas prone to landslides. Subsequent field expeditions validated potential landslide sites, assessed their dimensions and configurations, determined types of movement, conducted on-site inspections, and categorized their activity status (active, dormant, etc.).

For training, 43 specific locations were pinpointed and digitally mapped into raster format within a Geographic Information System (GIS), using a grid size of 30 meters by 30 meters. This grid size was chosen based on the Digital Elevation Model (DEM) extracted from the Shuttle Radar Topography Mission (SRTM) database, which was used to generate slope and elevation maps. Additional vector data layers such as lithology, curvature, and aspect were integrated into the analysis.

Stream maps were also rasterized using the same grid size. Machine learning models achieve optimal results when multiple relevant input variables are considered. Therefore, only significant variables with substantial impact on estimating the target value were included in the analysis [14].

Ten landslide-triggering parameters were selected for this study: lithology, elevation, slope, aspect, plan curvature, profile curvature, distance to streams, distance to roads, land cover, and rainfall. These parameters were chosen based on relevant literature, expert knowledge, and the specific characteristics of the study area.

- Elevation

  This factor exerts a significant influence on landslide occurrences, consequently impacting various other factors such as slope, erosion, precipitation, soil depth, and land use. Conducting a comprehensive analysis of the topographic relief is beneficial, enabling the identification of both the highest and lowest elevation areas within the terrain [15].The elevation ranges from 112 to 1292 meters (**Figure 3**) (**Table 1**).

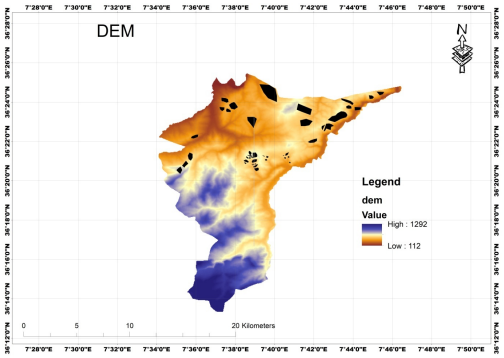

Figure 3. Digital elevation model.

- Lithology

  The study area features a variety of geological formations, each with unique characteristics that significantly influence landslide occurrence [16]. The geological data were extracted from the 1/200,000 scale geological map of Algeria. The geology has been divided into 6 classes, namely: marine Triassic, Pontian, marine Cretaceous, continental Eocene, marine Oligocene, and alluvium (**Figure 4**) (**Table 1**).

  These formations play a critical role in understanding and predicting landslides within the study area.

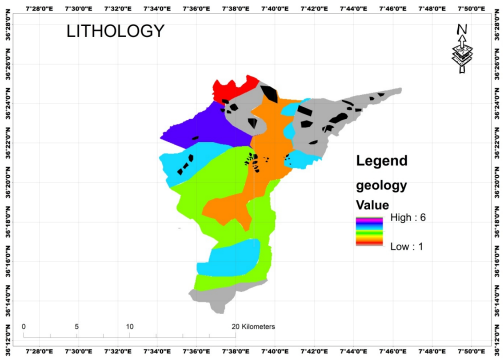

Figure 4. Lithology.

- Slope

  The slope gradient is widely recognized as the primary factor predisposing areas to landslides, influencing the accumulation of moisture and pore pressure locally. At a broader regional scale, it governs hydraulic connectivity and is a fundamental parameter for Geographic Information System (GIS)-based mapping. The slope angle closely correlates with landslide susceptibility, with steeper angles indicating higher vulnerability to landslides[17]. The slope angle ranges from 0 to 61.95 degrees (**Figure 5**) (**Table 1**).

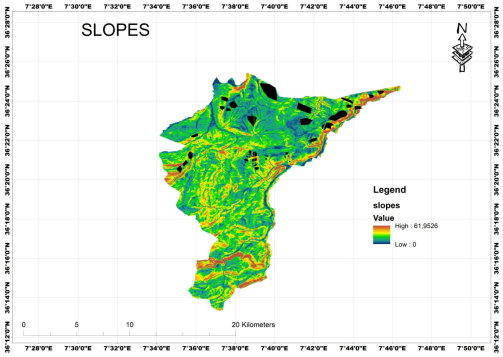

Figure 5. Slopes.

- Proximity of streams

  The proximity of a site to a nearby stream, known as "distance to stream," is a crucial factor in slope stability. Streams influence slope stability primarily through erosion and by saturating the lower part of the slope, which can lead to increased water levels and reduced soil mechanical properties[15]. The distance to streams ranges from 0 to 2881 meters (**Figure 6**) (**Table 1**).

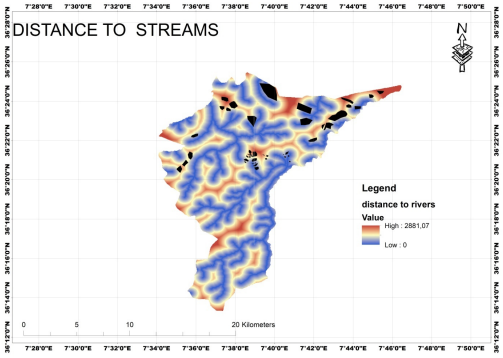

Figure 6. Distance to streams.

- Plan curvature

  Plan curvature, also known as cross-sectional curvature, measures the curvature of topographic contours or the curvature formed by the intersection of an imaginary horizontal plane with the ground surface. This metric quantifies curvature perpendicular to the direction of the steepest slope, providing insights into the horizontal configuration of the terrain. Positive plan curvature indicates convex shapes, whereas negative plan curvature suggests concave shapes. In landslide analysis, areas with negative plan curvature are typically considered more susceptible to landslides because they feature concave slopes that can collect water and sediments, thereby increasing instability risks[18]. In our case study, plan curvature ranges from −355364 to 318541 degrees (**Figure 7**) (**Table 1**).

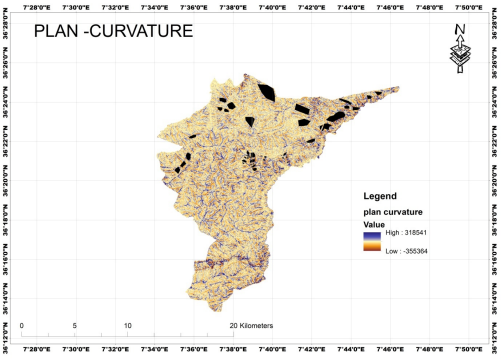

Figure 7. Plan curvature.

- Profile curvature

  Known as longitudinal curvature, this metric evaluates the change in slope along the path of the steepest descent, offering insights into whether the landscape is concave, convex, or flat. A notable positive profile curvature indicates ridges, while a significant negative profile curvature suggests valleys or depressions. Regions with pronounced negative profile curvature may exhibit increased susceptibility to landslides, as concave slopes can retain water and potentially facilitate instability. In our specific context[19], the profile curvature ranges from −643082 to 447453 degrees (**Figure 8**) (**Table 1**).

- Land cover

  Land cover exerts a significant influence on the susceptibility of an area to landslides, influenced by factors such as soil composition, vegetation type, and human activities. Various factors highlight the impact

of land cover on landslide vulnerability, including the density of vegetation, deforestation rates, urban development, soil degradation, water management practices, climatic factors, and land management strategies[20]. The study area was classified into eleven categories based on land cover: 1. Nodata, 2. Water bodies, 3. Areas covered by trees, 4. Grasslands, 5. Flooded vegetation, 6. Cultivated crops, 7. Scrublands, 8. Built-up areas, 9. Bare ground, 10. Snow-covered areas, 11. Clouds (**Figure 9**) (**Table 1**).

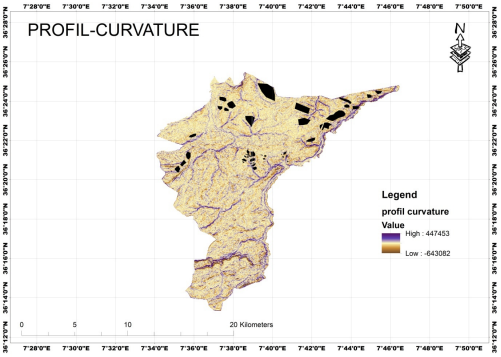

Figure 8. Profile curvature.

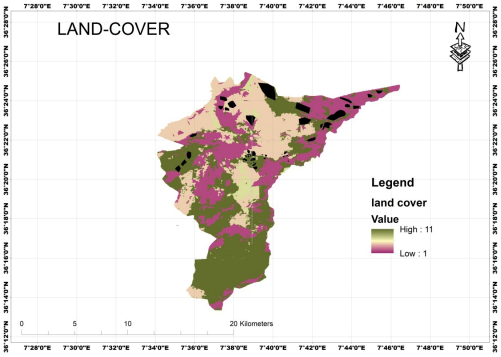

Figure 9. Land cover.

- Rainfall

  Precipitation plays a pivotal role in triggering landslides, particularly during periods of heavy rainfall .Rainfall induces several detrimental mechanisms that affect slope stability. Firstly, the infiltration of rainwater and subsequent soil saturation increase pore pressure, reducing the soil's shear strength. Additionally, runoff and soil erosion, both initiated by precipitation, further compromise soil stability. The accumulation of water within the ground can also lead to mudflows and mud avalanches, adding extra load and exacerbating stress within the soil, thereby increasing the likelihood of landslide occurrence[21]. In our case study, rainfall ranges from 550 to 600 mm annually (**Figure 10**) (**Table 1**).

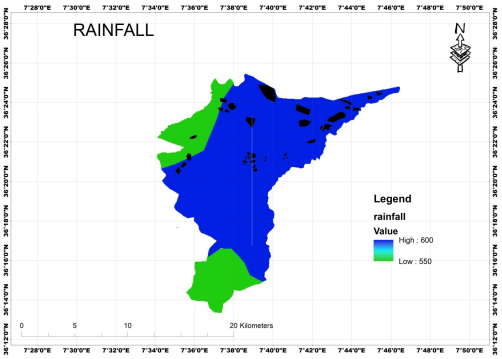

Figure 10. Rainfall.

- Aspect

  Slope aspect, indicating the compass direction a slope faces, is measured in degrees from north in a clockwise direction, ranging from 0° to 360°. This aspect significantly influences various environmental factors, such as sunlight exposure and prevailing wind direction. These factors, in turn, play a crucial role in shaping conditions that contribute to landslides, including precipitation patterns, snow accumulation, soil moisture levels, vegetation distribution, and soil depth[22].

  To demonstrate the influence of slope aspect, we created an aspect map of the study area, dividing it into nine distinct classes: flat, N (North), NE (Northeast), E (East), SE (Southeast), S (South), SW (Southwest), W (West), and NW (Northwest). Researchers, particularly in mid and high latitudes, generally agree that north-facing slopes (or south-facing in the southern hemisphere) and northwest-facing slopes are particularly prone to landslides. The aspect factor in our study ranges from −1 to 359.79 degrees (**Figure 11**) (**Table 1**).

- Proximity of roads

  Previous studies consistently highlight the proximity to roads as a significant factor influencing landslide occurrence. Especially in mountainous regions where roads often run alongside slopes, the excavation and construction processes can disrupt the natural soil balance, potentially triggering instability. Conversely, greater distances from roads reduce stress on the ter-

rain and slope bases, thereby decreasing landslide risk. Consequently, the layout of road networks plays a critical role in determining landslide susceptibility [23]. In this case study, distances to roads range from 0 to 11742 meters (**Figure 12**).

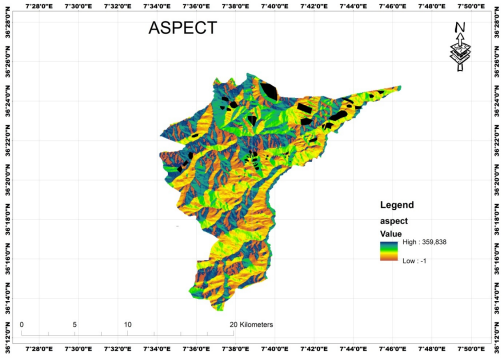

Figure 11. Aspect.

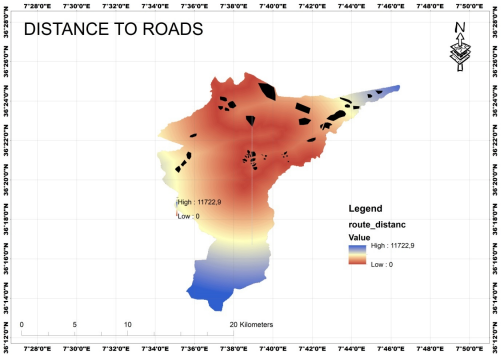

Figure 12. Distance to roads.

Table 1. Factors description.

Factor	Description-Value
Slope (º)	0−61.95
Lithology	marine Triassic, Pontian, marine Cretaceous, continental Eocene, marine Oligocene, alluvium.
Elevation (m)	112−1292
Distance to streams (m)	0−2881
Curvature profile (º)	−443082−447453
Curvature plan (º)	−355364−318541
Landcover	1—nodata, 2—water, 3—trees, 4—grass, 5—flooded vegetation, 6—crops, 7—scrub, 8—built, 9—bareground, 10—snow, 11—clouds
Rainfall (mm)	550−600
Aspect (º)	−1−359.79
Distance to roads (m)	0−11742

2.3. Inventory Map and Conditional Factor

Using Landsat imagery from Google Earth and field survey, there were 43 important landslides. The conditional factor is the delimitation of slopes to a maximum of 18 degrees (**Figure 13**).

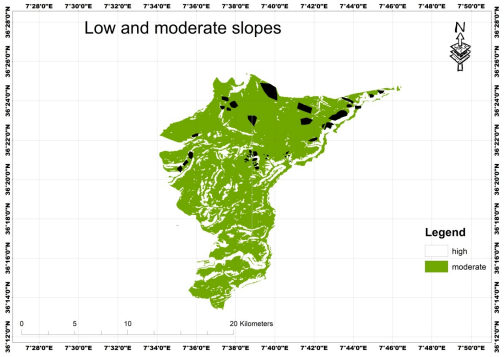

Figure 13. Moderate slopes map.

3. Methodology and Results

3.1. Bagging

Bagging, short for bootstrap aggregation, is a robust ensemble machine learning technique introduced by Breiman in 1996, (**Figure 14**). This method involves creating multiple subsets of the training data through a statistical resampling approach known as bootstrapping. By doing so, each subset becomes unique, fostering diversity within the ensemble In bagging, a base learner is trained on each bootstrap sample to generate individual predictions. These predictions are then combined using a majority voting strategy to determine the final prediction of the ensemble method. While various base learners can be employed [24, 25], decision trees are commonly favored due to their simplicity, high variance, and computational efficiency [26].

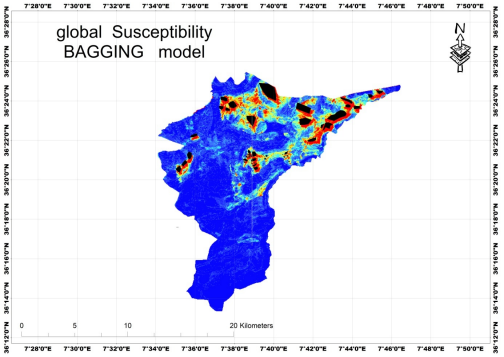

Figure 14. Global susceptibility—bagging model.

The bagging process is divided into steps with formulas:

1. Bootstrap Sampling

$$Dm = \{(x_i, y_i)\}_{i=1}^{N} \quad (1)$$

- For each base model m , a bootstrap sample Dm is drawn with replacement from the original dataset D
- The size of each bootstrap sample is the same as the original dataset, denoted as N
- The process is repeated M times to generate M bootstrap samples.

2. Model Training

$$f_m(\mathrm{x}) = \text{Model trained on Dm } (\mathrm{m} = 1, 2, 3, ..., \mathrm{m}) \quad (2)$$

3. Aggregation of Predictions (Classification):

$$Y = argmax \sum_{m=1}^{m} (f_m(x) = y) \quad (3)$$

For classification tasks, the final prediction for each instance x in the original dataset.

3.2. Multi Layer Perceptron (MLP)

A Multi-Layer Perceptron (MLP) is a type of artificial neural network (ANN) that consists of multiple layers of interconnected nodes, or neurons[27]. It is a versatile and powerful model used for supervised learning tasks such as classification and regression[28] (**Figure 15**). Here is a breakdown of the key components and functionalities of an MLP:

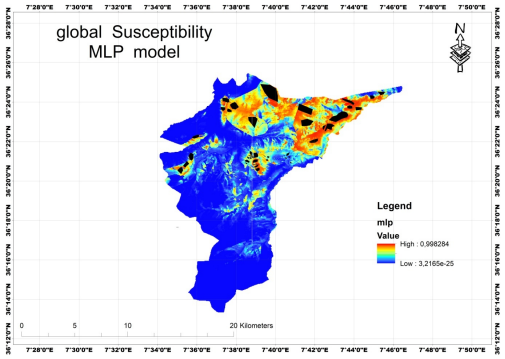

Figure 15. Global susceptibility—MLP model.

1. Input Layer

- The input layer comprises neurons that receive the initial data or features. Each neuron represents one feature of the input data.

2. Hidden Layers

- Between the input and output layers, there can be one or more hidden layers, each containing a certain number of neurons.
- Neurons in the hidden layers perform computations by taking weighted sums of inputs from the previous layer, followed by the application of an activation function. The activation function introduces non-linearity to the network, allowing it to learn complex patterns in the data .
- The number of hidden layers and neurons in each layer is a hyperparameter that needs to be determined based on the complexity of the problem and the available computational resources.

3. Output Layer

- The output layer produces the final predictions or outputs of the network. The number of neurons in the output layer depends on the type of task.
- For instance, in a binary classification task, there would typically be one output neuron representing the probability of belonging to one class. In a multi-class classification task, there would be multiple output neurons, each representing the probability of belonging to a different class.

4. Activation Functions

- Activation functions introduce non-linearity into the network, enabling it to learn complex relationships in the data.
- Common activation functions include sigmoid, tanh, and Rectified Linear Unit (ReLU) (Rumelhart et al., 1986). Each has its advantages and is suitable for different types of problems.

An MLP can learn and generalize from data through backpropagation, a method used for training the network by adjusting weights to minimize the error between predicted and actual outputs.

5. Training

- Training an MLP involves adjusting the weights and biases of the connections between neurons to minimize the difference between the predicted outputs and the actual targets.
- This is typically done using optimization algorithms such as gradient descent and computer vision backpropagation, which calculate the gradients of the loss function with respect to the weights and biases and update them accordingly (Rumelhart et al., 1986).

MLPs are widely used in various domains, including natural language processing, and financial forecasting, due to

their ability to model complex relationships in data and their scalability to large datasets. However, they may require careful tuning of hyperparameters and regularization techniques to prevent overfitting, especially for deep architectures with many layers.

Input function

$$j = \sum_{i=1}^{n} w_{ij}^{(1)} x_i + b_j^{(1)} \quad (4)$$

Activation function

Output of hidden node j = f(input to hidden node j)

Hidden to output layer

$$Input\ to\ output\ node\ k = \sum_{j=1}^{m} w_{jk}^{(2)} h_j + b_k^{(2)} \quad (5)$$

Output function

$$Output\ of\ output\ node\ k = g(input\ to\ output\ k) \quad (6)$$

Where

w_{ij} is the weight of the connection between input node i and hidden node j in the first layer; x_i is the input to input node i;

'f' is the activation function (sigmoid, tanh, relu);

$w_{jk}^{(2)}$ is the weight of the connection between hidden node j and out put node k in the second layer;

"g" is the activation function applied element wise; h_j is the output of hidden node 'j'; $b_k^{(2)}$ is the bias of output node k.

3.3. Random Forest (RF)

Random forest, an ensemble learning algorithm introduced by Breiman in 2001, stands out as one of the most powerful methods in machine learning (**Figure 16**). As the name suggests, random forest employs multiple decision trees as base learners, each trained on bootstrap samples derived from the training dataset. What sets random forest apart from basic bagging techniques is its method of introducing diversity among trees. At each split within a tree, a random subset of features from the full set is considered as candidates for the split, enhancing the algorithm's robustness. Prediction in random forest is based on a voting process among the ensemble of trees.

This algorithm has become a cornerstone in the field of landslide susceptibility mapping, demonstrating remarkable success[29]. Balancing performance and computational efficiency, random forest requires minimal hyperparameter tuning to yield effective models.

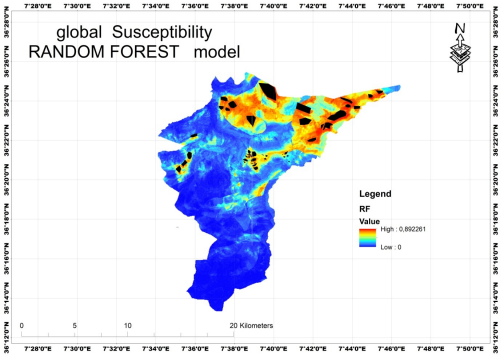

Figure 16. Global susceptibility—random forest (RF) model.

1. Bootstrap Sampling

$$Dm = \{(x_i, y_i)\}_{i=1}^{N}, i = 1, 2, 3, 4, ..., B \quad (7)$$

2. Feature Subset Selection:

At each node of each decision tree, a random subset of features is selected for splitting. This ensures diversity among the trees.

3. Decision tree training

For each bootstrap sample, a decision tree Ti is grown using a subset of features selected at random at each node

4. Prediction aggregation—classification

$$Y = argmax \sum_{i=1}^{B} (T_i(x) = y) \quad (8)$$

3.4. Support Vector Machine (SVM)

Initially, Support Vector Machines (SVM) serve as binary classifiers, assigning instances to one of two classes (**Figure 17**). However, they can be easily adapted for multiclass scenarios by transforming them into a series of binary tasks using either one-versus-all or one-versus-one approaches[30].

The core concept of the SVM algorithm involves establishing a hyperplane in the original n-dimensional space (with xi parameters in vector x) to separate points belonging to different classes. What distinguishes SVM from other methods is how this hyperplane is derived from the training set. The algorithm aims to maximize the margin between classes (M2 NM1) and positions a classification hyperplane at the midpoint of this margin (f(x)). Points in Rn̂ above the hyperplane are labeled as +1, while those below are labeled

as −1. The classification function is expressed by Equation (1), where y denotes the class label, w and b are hyperplane parameters, and sgn is the sign function. Studies have demonstrated that a maximum-margin classifier exhibits superior generalization performance on unseen data compared to alternative hyperplane solutions (8).

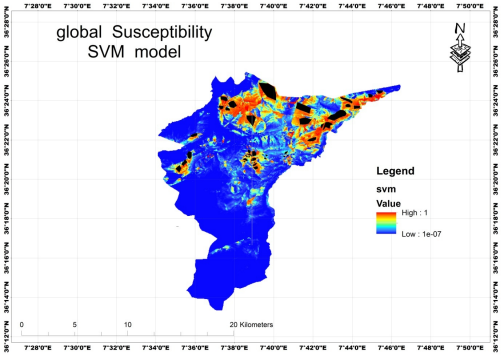

Figure 17. Global susceptibility—support vector machine (SVM) model.

Real-world classification datasets often contain noise and non-linear separability, allowing some points to fall on the incorrect side of the hyperplane. Consequently, a soft-margin classifier is designed to strike a balance between generalization capacity (margin width) and empirical error (sum of training errors ek). Adjusting a parameter C during SVM training enables control over this trade-off; higher C values yield more intricate models.

The optimal hyperplane weights w are determined as a linear combination of training points, expressed as

$$y = sgn\left(f\left(x\right)\right) = sgn\left(\sum_{i=1}^{n} \omega_i.x_i + b\right) = sgn(wx+b) \quad (9)$$

$$f\left(x\right) = sgn\sum_{i=1}^{n} \varphi_i yi\left(x_i.x\right) + b \quad (10)$$

3.5. K-Nearest Neighbors (KNN)

nearest neighbors (KNN) is a machine learning algorithm often employed for predicting landslides[31, 32], (**Figure 18**).

Here's a description of how KNN can be used for landslide prediction:

1. Data Collection: Relevant data on factors influencing landslides is gathered. This might include information on terrain, soil type, land use, precipitation patterns, geological features, vegetation cover, and historical landslide occurrences.

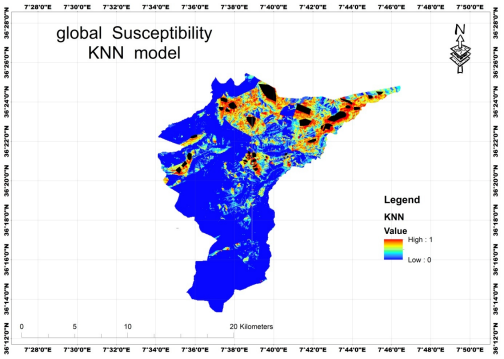

Figure 18. Global susceptibility—K-nearest neighbors (KNN) model.

2. Data Preprocessing: The collected data is cleaned, processed, and formatted for analysis. This involves handling missing values, normalizing or scaling features, and ensuring data consistency.

3. Feature Selection: Key features contributing to landslide occurrences are identified. This step involves analyzing the collected data and selecting the most relevant features that have a significant impact on landslide susceptibility.

4. Training Data Preparation: The dataset is divided into two parts: a training set and a testing set. The training set is used to train the KNN model, while the testing set is used to evaluate its performance.

5. Model Training: The KNN model is trained using the training dataset. During training, the model learns the relationships between the selected features and landslide occurrences by storing the feature vectors of training data points.

6. Choosing K Value: The value of K (the number of nearest neighbors) is chosen. This parameter affects the model's performance and generalization ability. Typically, the optimal K value is determined through experimentation and validation.

7. Prediction: To predict whether a given location is susceptible to landslides, the KNN algorithm calculates the distances between the feature vector of the target location and the feature vectors of its K-nearest neighbors in the training dataset.

8. Majority Voting (Classification): For landslide classification, the algorithm performs a majority voting among the K-nearest neighbors to determine the landslide susceptibility of the target location. The class with the highest

number of votes among the neighbors is assigned to the target location.

Here are formulas used in the K-nearest neighbors (KNN) algorithm:

1. Euclidean Distance: The Euclidean distance is commonly used to measure the distance between two data points x_i and x_j in a feature space. It is calculated as:

$$\text{Euclidean distance} = \sqrt{\sum_{k=1}^{n} (x_{ik} - x_{jk})^2} \quad (11)$$

2. KNN Classification: In KNN classification, the class of a new data point x is determined by the majority class among its K-nearest neighbors. The predicted class y is calculated as:

$$Y = argmax \sum_{i=1}^{k} (I(Y_i) = j) \quad (12)$$

I is indicator function i = 1 if the condition inside is true, and = 0 otherwise.

3.6. A Decision Tree (DT) Model

A decision tree (DT) model, characterized by its hierarchical tree structure, serves as a non-parametric method adept at uncovering non-linear and non-additive relationships between input factors and predictive variables [33]. Described as a pattern recognition tool, DTs delineate structural patterns in data through a series of rules [34]. These rules can encompass various types of factors, including binary, nominal, ordinal, and quantitative, while the classes are qualitative in nature [35] (**Figure 19**).

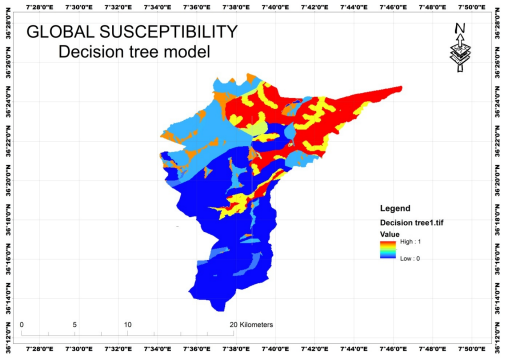

Figure 19. Global susceptibility—decision tree model.

For a given dataset comprising factors and their corresponding classes, a DT generates a sequence of rules to classify unseen records. The model's structure and decision-making process are illustrated in a tree format, consisting of a root node representing the primary decision point based on the most influential predictor variable, internal nodes denoting decision points with multiple branches, and terminal nodes representing final classifications or decisions [36]. Each node in the decision tree makes a binary decision that partitions classes, with traversal occurring down the tree until a terminal node is reached.

Several algorithms for decision tree learning exist, such as Classification and Regression Trees (CART), Iterative Dichotomiser 3 (ID3) [37], and C4.5 [38]. These algorithms differ in their approach to quantifying distinction and diversity criteria but share the hypothesis that entities within the dataset may be independent.

The fundamental algorithm for constructing decision trees, ID3, was developed by Quinlan in the mid-1980s. ID3 follows a top-down approach, utilizing entropy and information gain as criteria to guide the search process [37]. In information theory, entropy, denoted as H(D), measures uncertainty within a dataset. Information gain, represented as (Gain(D,A)), quantifies the reduction in entropy at subsequent hierarchy levels, where datasets are refined using supporting attributes [39].

$$H(D) = -\sum_{I=1}^{n} p_i log_2 p_i \quad (13)$$

$$Gain(D, A) = H(D) - \sum_{v \in ValueA} \frac{Dv}{D} H(Dv) \quad (14)$$

3.7. Application to the Specific Case of Low and Moderate Slopes

The susceptibility to landslides of slopes with low and moderate gradients results, for each model, from the overlay of the global susceptibility maps with the map of low and moderate slopes using the Raster Calculator tools in ArcGIS (**Figures 20–24**).

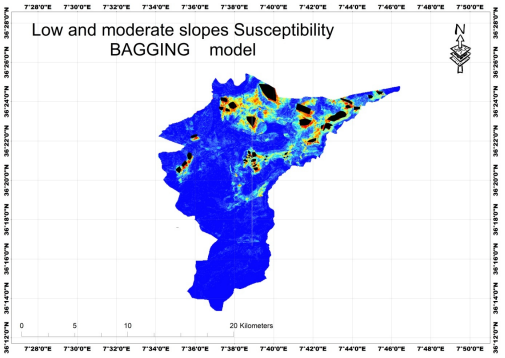

Figure 20. Low slopes susceptibility—bagging model.

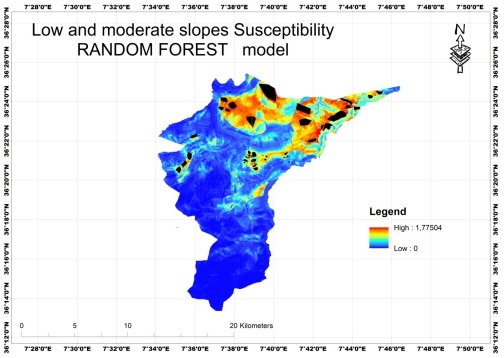

Figure 21. Low slopes susceptibility—random forest model.

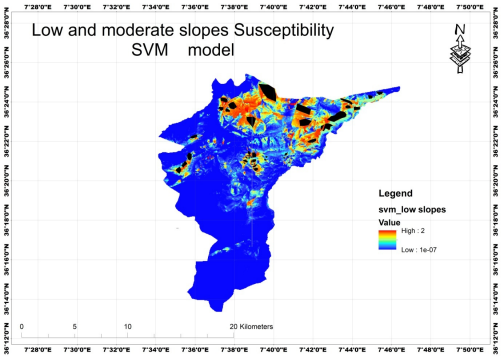

Figure 22. Low slopes susceptibility—SVM model.

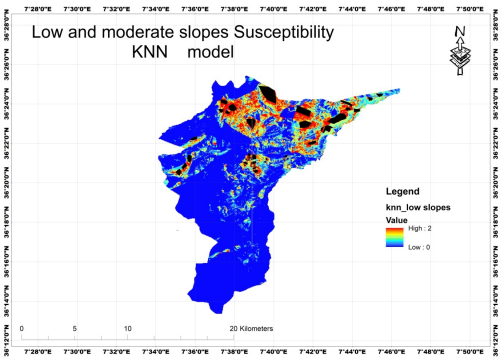

Figure 23. Low slopes susceptibility—KNN model.

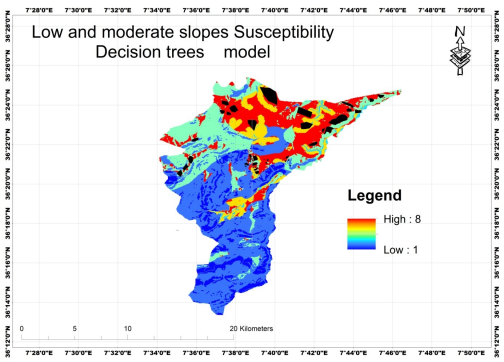

Figure 24. Low slopes susceptibility—decision trees model.

4. Discussion

The susceptibility to landslides affecting the low and medium slopes of the Hamam Nbail site is estimated by several algorithms using Python, including Random Forest, Bagging, Decision Tree, Support Vector Machine, K-Nearest Neighbors, and Machine Learning Perceptron. All of them have provided acceptable performances. The methods adopted to estimate the performances include the ROC_AUC curve, as explicitly shown in the figures below (**Figures 25–30**).

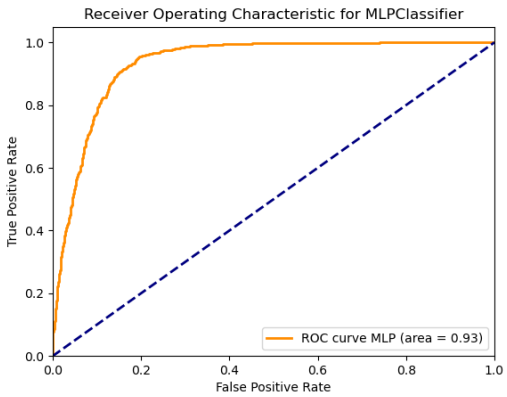

Figure 25. ROC_AUC curve—MLP.

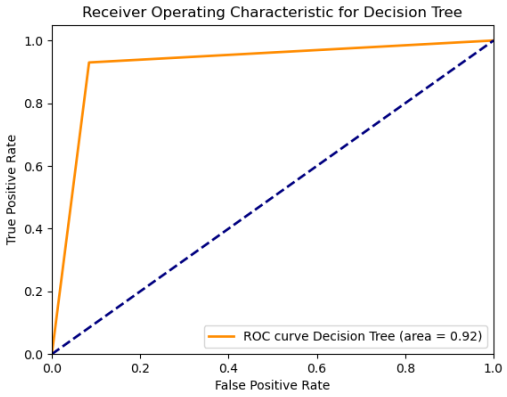

Figure 26. ROC_AUC curve—DT.

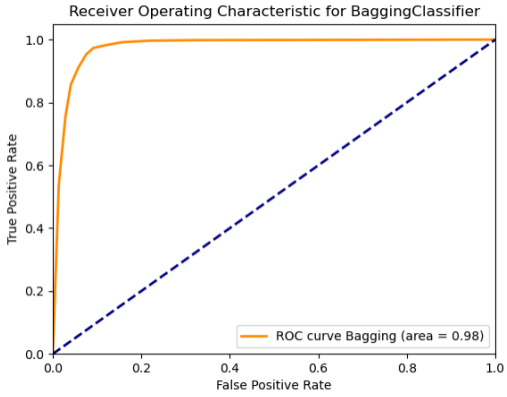

Figure 27. ROC_AUC curve—bagging.

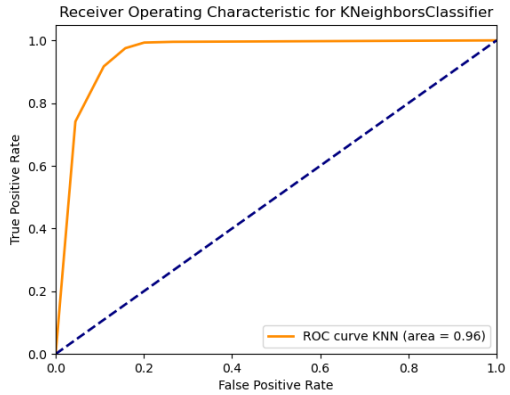

Figure 28. ROC_AUC curve—KNN.

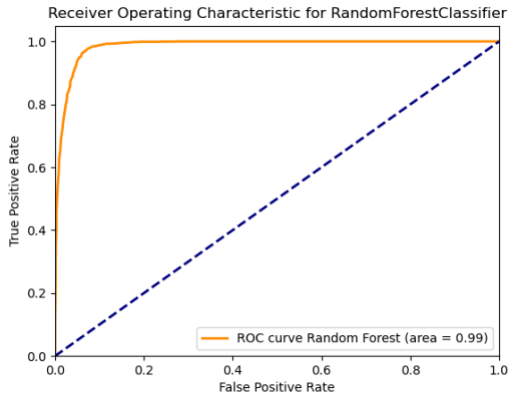

Figure 29. ROC_AUC curve —RF.

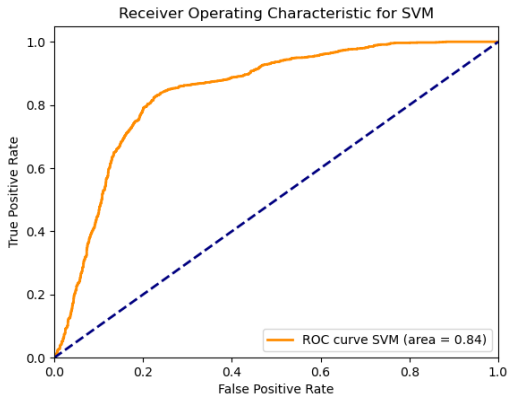

Figure 30. ROC_AUC curve—SVM.

5. Conclusions

As the completion of the study on landslide susceptibility for the low and medium slope hillsides of the Hamam Nbail site in eastern Algeria approaches, it is evident that among the proposed models, the best performances were achieved using the Random Forest algorithm (**Figure 31**), whether in terms of ROC_AUC. Determining susceptibility for these low and medium slopes is crucial as significant infrastructure development is planned for these areas. Hence, understanding the vulnerability of these terrains to landslides is necessary (**Figure 31**).

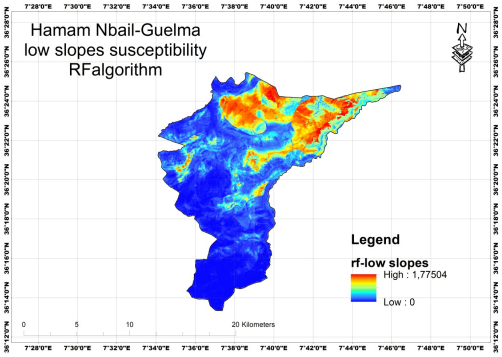

Figure 31. Final low slopes susceptibility map using random forest.

Author Contributions

Conceptualization: H.A.; Methodology: D.L.; Supervision: B.R.

Funding

This work received no external funding.

Institutional Review Board Statement

This study did not require approval from an ethics committee, as it does not involve data collection from human or animal participants, nor does it include elements subject to specific ethical considerations.

Informed Consent Statement

Not applicable for this study, not involving humans.

Data Availability Statement

The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Conflicts of Interest

The authors declare no conflict of interest for this article.

References

[1] Djerbal, L., Melbouci, B., 2012. Le glissement de terrain d'Ain El Hammam (Algerie): causes et evolution. Bull Eng Geol Environ. 71(3). Available from: https://search.ebscohost.com/login.aspx?direct=true&profile=ehost&scope=site&authtype=crawler&jrnl=14359529&AN=78312659&h=eM8K%2F60uULXGO639j2P6lt1VevXgptFYtNOX%2B2CvaATW9R6nqX2pIYDH4Y1xDDNP7L%2BaS99VclKRmBorAIOjeg%3D%3D&crl=c (cited 27 October 2024).

[2] Achour, Y., Saidani, Z., Touati, R., et al., 2021. Assessing landslide susceptibility using a machine learning-based approach to achieving land degradation neutrality. Environmental Earth Sciences. 80(17), 575. DOI: https://doi.org/10.1007/s12665-021-09889-9

[3] Kadavi, P.R., Lee, C.W., Lee, S., 2019. Landslide-susceptibility mapping in Gangwon-do, South Korea, using logistic regression and decision tree models. Environmental Earth Sciences. 78(4), 116. DOI: https://doi.org/10.1007/s12665-019-8119-1

[4] Nourani, V., Pradhan, B., Ghaffari, H., et al., 2014. Landslide susceptibility mapping at Zonouz Plain, Iran using genetic programming and comparison with frequency ratio, logistic regression, and artificial neural network models. Nat Hazards. 71(1), 523–547. DOI: https://doi.org/10.1007/s11069-013-0932-3

[5] Nhu, V.H., Shirzadi, A., Shahabi, H., et al., 2020. Shallow landslide susceptibility mapping: A comparison between logistic model tree, logistic regression, naïve bayes tree, artificial neural network, and support vector machine algorithms. International Journal of Environmental Research and Public Health. 17(8), 2749. DOI: https://doi.org/10.3390/ijerph17082749

[6] Akgun, A., 2012. A comparison of landslide susceptibility maps produced by logistic regression, multi-criteria decision, and likelihood ratio methods: a case study at İzmir, Turkey. Landslides. 9(1), 93–106. DOI: https://doi.org/10.1007/s10346-011-0283-7

[7] Akgun, A., Türk, N., 2010. Landslide susceptibility mapping for Ayvalik (Western Turkey) and its vicinity by multicriteria decision analysis. Environmental Earth Sciences. 61(3), 595–611. DOI: https://doi.org/10.1007/s12665-009-0373-1

[8] Lee, S., Hong, S.M., Jung, H.S., 2017. A support vector machine for landslide susceptibility mapping in Gangwon Province, Korea. Sustainability. 9(1), 48. DOI: https://doi.org/10.3390/su9010048

[9] Ohlmacher, G.C., Davis, J.C., 2003. Using multiple logistic regression and GIS technology to predict landslide hazard in northeast Kansas, USA. Engineering Geology. 69(3–4), 331–343. DOI: https://doi.org/10.1016/S0013-7952(03)00069-3

[10] Didier, C., Tritsch, J.J., Watelet, J.M., et al., 1999. Évaluation du risque d'instabilité en surface à l'aplomb d'une ancienne carrière souterraine. Principes d'une analyse par configurations types. In: 9 Congrès International de Mécanique des Roches. pp. 9–14. Available from: https://ineris.hal.science/ineris-00972172/ (cited 27 October 2024).

[11] Pradhan, B., 2010. Application of an advanced fuzzy logic model for landslide susceptibility analysis. International Journal of Computational Intelligence Systems. 3(3), 370. DOI: https://doi.org/10.2991/ijcis.2010.3.3.12

[12] Kavzoglu, T., Sahin, E.K., Colkesen, I., 2014. Landslide susceptibility mapping using GIS-based multi-criteria decision analysis, support vector machines, and logistic regression. Landslides. 11(3), 425–39. DOI: https://doi.org/10.1007/s10346-013-0391-7

[13] Ayalew, L., Yamagishi, H., 2005. The application of GIS-based logistic regression for landslide susceptibility mapping in the Kakuda-Yahiko Mountains, Central Japan. Geomorphology. 65(1–2), 15–31. DOI: https://doi.org/10.1016/j.geomorph.2004.06.010

[14] Ayalew, L., Yamagishi, H., Marui, H., et al., 2005. Landslides in Sado Island of Japan: Part II. GIS-based susceptibility mapping with comparisons of results from two methods and verifications. Engineering Geology. 81(4), 432–445. DOI: https://doi.org/10.1016/j.enggeo.2005.08.004

[15] Achour, Y., Pourghasemi, H.R., 2020. How do machine learning techniques help in increasing accuracy of landslide susceptibility maps? Geoscience Frontiers. 11(3), 871–883. DOI: https://doi.org/10.1016/j.gsf.2019.10.001

[16] Van Westen, C.J., Van Asch, T.W.J., Soeters, R., 2006. Landslide hazard and risk zonation—why is it still so difficult? Bull Eng Geol Environ. 65(2), 167–84. DOI: https://doi.org/10.1007/s10064-005-0023-0

[17] Sharma, R.H., 2013. Evaluating the effect of slope curvature on slope stability by a numerical analysis. Australian Journal of Earth Sciences. 60(2), 283–290. DOI: https://doi.org/10.1080/08120099.2013.762942

[18] Ohlmacher, G.C., 2007. Plan curvature and landslide probability in regions dominated by earth flows and earth slides. Engineering Geology. 91(2–4), 117–134. DOI: https://doi.org/10.1016/j.enggeo.2007.01.005

[19] Keller, E.A., Bean, G., Best, D., 2015. Fluvial geomorphology of a boulder-bed, debris-flow—Dominated channel in an active tectonic environment. Geomorphology. 243, 14–26. DOI: https://doi.org/10.1016/j.geomorph.2015.04.012

[20] Promper, C., Puissant, A., Malet, J.P., et al., 2014. Analysis of land cover changes in the past and the future as contribution to landslide risk scenarios. Applied Geography. 53, 11–19. DOI: https://doi.org/10.1016/j.apgeog.2014.05.020

[21] Matougui, Z., Djerbal, L., Bahar, R., 2023. A

comparative study of heterogeneous and homogeneous ensemble approaches for landslide susceptibility assessment in the Djebahia region, Algeria. Environ Sci Pollut Res. 31(28), 40554–40580. DOI: https://doi.org/10.1007/s11356-023-26247-3

[22] Cellek, S., 2021. The effect of aspect on landslide and its relationship with other parameters. In: Landslides. IntechOpen. Available from: https://books.google.com/books?hl=fr&lr=&id=LLGTEAAAQBAJ&oi=fnd&pg=PA13&dq=aspect+and+landslides&ots=GgtlXVrNhN&sig=EqWdKBtrDlcb3HmcKdsMkUb7Iys (cited 27 October 2024).

[23] Çellek, S., 2016. Linear parameters causing landslides: A case study of distance to the road, fault, drainage. Kocaeli Journal of Science and Engineering. 6(2), 94–113. DOI: https://doi.org/10.34088/kojose.1117817

[24] Pham, B.T., Tien, Bui, D., Prakash, I., 2017. Landslide Susceptibility Assessment Using Bagging Ensemble Based Alternating Decision Trees, Logistic Regression and J48 Decision Trees Methods: A Comparative Study. Geotech Geol Eng. 35(6), 2597–611. DOI: https://doi.org/10.1007/s10706-017-0264-2

[25] Dou, J., Yunus, A.P., Bui, D.T., et al., 2020. Improved landslide assessment using support vector machine with bagging, boosting, and stacking ensemble machine learning framework in a mountainous watershed, Japan. Landslides. 17(3), 641–658. DOI: https://doi.org/10.1007/s10346-019-01286-5

[26] Zhao, Y., Zhang, Y., 2008. Comparison of decision tree methods for finding active objects. Advances in Space Research. 41(12), 1955–1959. DOI: https://doi.org/10.1016/j.asr.2007.07.020

[27] Bishop, C.M., 1995. Neural networks for pattern recognition. Oxford University Press. Available from: https://books.google.com/books?hl=fr&lr=&id=T0S0BgAAQBAJ&oi=fnd&pg=PP1&dq=Bishop,+1995&ots=jO5TqJ6Bli&sig=mduOPrqben_hr8wa7lsPFrQ9IyI (cited 27 October 2024).

[28] Goodfellow, I., Bengio, Y., Courville, A., 2016. Deep feedforward networks. Deep Learn. Available from: https://mnassar.github.io/deeplearninghandbook/slides/06_mlp.pdf (cited 27 October 2024).

[29] Cheng, Y.S., Yu, T.T., Son, N.T., 2021. Random forests for landslide prediction in Tsengwen river watershed, central Taiwan. Remote Sensing. 13(2), 199. DOI: https://doi.org/10.3390/rs13020199

[30] Raschka, S., 2014. An Overview of General Performance Metrics of Binary Classifier Systems. arXiv. Available from: http://arxiv.org/abs/1410.5330 (cited 27 October 2024).

[31] Zhang, Z., 2016. Introduction to machine learning: k-nearest neighbors. Annals of Translational Medicine. 4(11), 218. DOI: https://doi.org/10.21037/atm.2016.03.37

[32] Pandey, A., Jain, A., 2017. Comparative analysis of KNN algorithm using various normalization techniques. International Journal of Computer Network and Information Security. 10(11), 36. Available from: https://www.researchgate.net/profile/Achin-Jain-2/publication/320932252_Comparative_Analysis_of_KNN_Algorithm_using_Various_Normalization_Techniques/links/652769a261c4044c404cda3a/Comparative-Analysis-of-KNN-Algorithm-using-Various-Normalization-Techniques.pdf (cited 27 October 2024).

[33] Bell, D.A., 1978. Decision Trees, Tables, and Lattices. In: Batchelor, B.G., ed. Pattern Recognition. Springer US. pp. 117–141. DOI: https://doi.org/10.1007/978-1-4613-4154-3_5

[34] Witten, I.H., Frank, E., Hall, M.A., et al., 2005. Practical machine learning tools and techniques. In: Data mining. Elsevier, Amsterdam, The Netherlands. pp. 403–413. Available from: https://sisis.rz.htw-berlin.de/inh2012/12401301.pdf (cited 27 October 2024).

[35] Murthy, S.K., Salsberg, S.L., 1998. Investigations of the greedy heuristic for classification tree induction. Available from: https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=e212e4a7609b732125fa011b2476d3e27abb65b8 (cited 27 October 2024).

[36] Breiman, L., Ihaka, R., 1984. Nonlinear discriminant analysis via scaling and ACE. Department of Statistics, University of California Davis One Shields Avenue. Available from: https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=ea4bdeb0ba42ed22c8529052837294167868075 5 (cited 27 October 2024).

[37] Quinlan, J.R., 1987. Generating production rules from decision trees. In: ijcai. Citeseer. pp. 304–307. Available from: https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=ea4bdeb0ba42ed22c8529052837294167868075 5 (cited 27 October 2024).

[38] Hssina, B., Merbouha, A., Ezzikouri, H., et al., 2014. A comparative study of decision tree ID3 and C4.5. Int J Adv Comput Sci Appl. 4(2), 13–19. Available from: https://www.academia.edu/download/34582216/Paper_3-A_comparative_study_of_decision_tree_ID3_and_C4.5.pdf (cited 27 October 2024).

[39] Li, X., Claramunt, C., 2006. A Spatial Entropy-Based Decision Tree for Classification of Geographical Information. Trans GIS. 10(3), 451–467. Available from: https://onlinelibrary.wiley.com/doi/10.1111/j.1467-9671.2006.01006.x (cited 27 October 2024).