Vietnam Journal of Earth Sciences, 48(1), 144−166, https://doi.org/10.15625/2615-9783/24023

Vietnam Journal of Earth Sciences
https://vjs.ac.vn/index.php/jse

Landslide susceptibility mapping using Partial Decision Tree-Based hybrid artificial intelligence models

Tran Van Phong[1,2], Pham The Truyen[1,2], Nguyen Van Duong[1,2], Nguyen Le Minh[1,2], Nguyen Dang Mau[3], Indra Prakash[4], Dao Minh Duc[1], Dam Nguyen Duc[5,*]

[1]*Institute of Earth Sciences, VAST, Hanoi, Vietnam*
[2]*Graduate University of Science and Technology, VAST, Hanoi, Vietnam*
[3]*Vietnam Institute of Meteorology, Hydrology and Climate Change, Hanoi, Vietnam*
[4]*DDG (R) Geological Survey of India, Gandhinagar 382010, India*
[5]*Geotechnical and Artificial Intelligence research group, University of Transport and Technology, Hanoi, Vietnam*

Received 07 November 2025; Received in revised form 12 December 2025; Accepted 29 December 2025

ABSTRACT

In this research, two newly hybrid machine learning (ML) models, including Decorate Ensemble-based Partial Decision Trees (D-PART) and Bagging Ensemble-based Partial Decision Trees (B-PART), were applied to generate an accurate landslide susceptibility map for the Muong Te area, Lai Chau Province, Vietnam. The performance of the novel models was compared with two single benchmark models, namely Partial Decision Trees (PART) and Logistic Regression (LR), using the popular area under the Receiver Operating Characteristic (ROC) curve (AUC) metric. To construct the training and validation datasets, a spatial database was developed comprising ten landslide conditioning factors associated with the area's topographic, geological, structural, and hydrological characteristics, along with 248 documented historical and recent landslide occurrences. The OneR technique was applied to prioritize the most influential factors and to improve the model's performance. The evaluation results demonstrate that D-PART yielded the strongest predictive performance, with an AUC of 0.801, followed by B-PART (0.795), PART (0.758), and Logistic Regression (0.736). Thus, the novel hybrid model D-PART is a promising technique for constructing a reliable landslide susceptibility map, which can be used for effective planning and management of landslides in landslide-prone areas.

Keywords: Machine learning, landslide susceptibility, ensemble models, Partial Decision Trees, Decorate, PART, Vietnam.

1. Introduction

Under the influence of gravity, rock masses, soil, and other materials may move downslope, a process commonly referred to as a landslide (Anbalagan, 1992), resulting from both natural and anthropogenic factors (Sameen et al., 2020; Prakash and Pham, 2024). Landslides cause substantial loss of life and severe damage to property, infrastructure, and disrupt communication all over the world. Incidents of landslides have been increasing over the last few decades due to development activities and the effects of climate change. Accurately predicting the timing of landslide occurrences remains highly challenging.

*Corresponding author, Email: damnd@utt.edu.vn

However, areas with potential future landslide occurrences can be identified and mapped through systematic and comprehensive analyses. By implementing appropriate management strategies and methods, it is thus possible to prevent landslides or reduce their effects (Nohani et al., 2019).

Various techniques have been used for the assessment of the susceptibility of landslides in different countries, such as Algeria (Achour et al., 2017; Achour and Pourghasemi, 2020; Achour et al., 2021), Slovakia (Pham et al., 2021b), and China (Chen et al., 2018), which can be broadly classified into quantitative and qualitative approaches. Qualitative (knowledge-based) approaches, exemplified by the Analytic Hierarchy Process (Anis et al., 2019), and fuzzy hierarchical analysis, rely primarily on expert judgment and are consequently classified as subjective methods (Yang et al., 2017b; Feizizadeh et al., 2014). Quantitative methods employ numerical techniques to evaluate the mathematical relationships between past landslide events and influencing factors; therefore, they are commonly referred to as objective approaches. Within quantitative frameworks, two main categories are recognized: deterministic and statistical methods. Most of these deterministic methods are effective but not very accurate at predicting landslides, as they do not account for the spatial relationships among factors across large areas. Conversely, statistical approaches such as the Evidential Belief Function, Frequency Ratio, Weight of Evidence (Bordbar et al., 2022; Shahabi and Hashim, 2015), and logistic regression (Nhu et al., 2020) evaluate landslide susceptibility by examining the spatial relationships between conditioning factors and historical landslide occurrences. In recent years, numerous studies have employed advanced machine learning (ML) techniques to explore landslide susceptibility, including Adaptive NeuroFuzzy Inference System (ANFIS), Incremental Generalized Model (IGM), and Support Vector Machine (SVM) (Chen et al., 2017a). Another investigation employed alternative machine learning algorithms, including Random Forest (RF), Logistic Model Tree (LMT), and Classification And Regression Tree (CART), for landslide susceptibility zoning and validated them as excellent techniques for landslide susceptibility mapping (Chen et al., 2017c).

The primary objective of this study is to assess the performance of two newly proposed ensemble-based hybrid machine learning models: Decorate ensemble-based Partial Decision Trees (D-PART) and Bagging ensemble-based Partial Decision Trees. (B-PART) for mapping landslide susceptibility in Muong Te area, Lai Chau, Vietnam. In addition, two benchmark single models, namely Partial Decision Trees (PART) and Logistic Regression (LR), are implemented for comparative assessment. This study introduces, for the first time, the use of D-PART and B-PART models for landslide susceptibility analysis in the study area, integrating a OneR-based feature selection framework. The study further evaluates model robustness using multiple statistical indicators, including AUC, RMSE, accuracy, sensitivity, and specificity, thereby providing a comprehensive and reliable assessment framework.

2. Study area

Covering an area of about 3,662 km², the study region is located in the Muong Te area of Lai Chau Province, Vietnam (Fig. 1). It is mainly mountainous terrain traversed by the Da River and its tributaries. The area is tectonically active and characterized by complex geological structures, including the Indosinian Fold Belt and the Song Da Rift System. The Indosinian Fold Belt comprises a

complex system of folded and faulted rock units that developed from the Late Paleozoic to the Early Mesozoic. These rocks consist mainly of metamorphic and sedimentary rocks, including schists, gneisses, phyllites, quartzites, limestones, and sandstones. The Song Da Rift System comprises a series of rift basins formed during the Late Mesozoic to Cenozoic period and characterized by thick sedimentary sequences. The regional rift system includes the Red River Fault, a major strike-slip structure extending along the eastern margin of the province and a key driver of the area's tectonic activity. The complex geological structure of Lai Chau Province has resulted in a diverse landscape featuring high mountains, deep valleys, and extensive river systems.

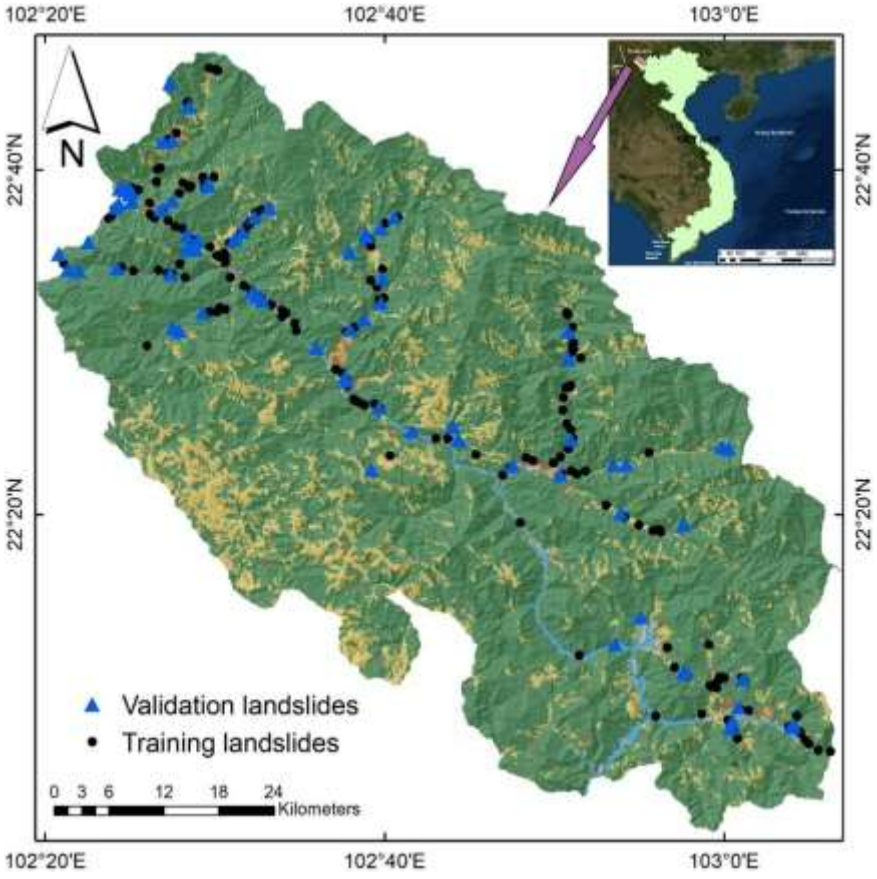

Figure 1. Geographic location of the study area in Lai Chau Province, Vietnam

The climate of this area can be grouped into two distinct seasons: Winter: cold, with little rain; Summer: hot, humid, rainy. Every year, the rainy season runs from the end of April to October, coinciding with the prevailing southwest wind. Rainfall in the highlands is up to 3000 mm/year, while the average across the mountain range ranges from 2000 to 2500 mm. Low-lying mountains and valleys receive rainfall of 1500 to 1800 mm. The dry season lasts from November to March, with a small amount of precipitation (316.8 mm) and long periods of fog, which are likely to occur during this time. The mean annual rainfall in the study area is approximately 2,531 mm, with precipitation

peaking in July and accounting for about 87.5% of the total. The average yearly temperature is around 22°C.

3. Materials and methods

This study followed a multi-step methodological framework. Firstly, a landslide inventory of 248 past and present landslide locations was collected. In addition, as landslide susceptibility modeling represents a binary classification task, 248 non-landslide samples were identified to balance the dataset. Of the total dataset, 70% of the landslide and non-landslide samples were randomly selected to construct the training dataset, while the remaining 30% were reserved for validation. The choice of a 70/30 split between the training and validation datasets was guided by the authors' experience and supported by findings reported in the literature (Asadi Nalivan et al., 2022; Mehrabi and Moayedi). In the subsequent stage, ten thematic maps representing landslide conditioning factors were constructed to examine their spatial relationships with landslide occurrences in the study area. The OneR feature selection algorithm was then applied to evaluate and identify the most influential factors for model development. Based on the selected features, two hybrid models D-PART and B-PART, and two single benchmark models, PART and Logistic Regression (LR), were developed for landslide susceptibility modeling. In this framework, D-PART integrates the Decorate ensemble technique with the PART classifier, whereas B-PART combines Bagging with PART. The performance of all models was subsequently assessed and compared using multiple validation metrics, including AUC, ACC, SST, SPE, PPV, NPV, and RMSE. Finally, landslide susceptibility maps were generated based on the outputs of the developed models. All machine learning models were implemented using the WEKA machine learning software, and model hyperparameters were optimized using a trial-and-error approach based on predictive performance and guidance from previous studies.

3.1. Data used

3.1.1. Spatial locations of past landslides and non-landslides

Ideal landslide inventory maps provide essential details on the landslide location, type, mode of movement, and underlying causes and triggering factors (e.g., earthquakes, rapid snowmelt, and heavy rainfall) (Althuwaynee et al., 2014). However, in the present study, only the spatial locations of historical landslides were utilized to compile the landslide inventory for spatial prediction within the study area. A total of 248 historical and recent landslide locations were identified through interpretation of Google Earth Pro imagery and field surveys conducted under the Vietnam national project (code: 01/2021/DX), entitled "Assessing the Causes of Earthquakes in Muong Te on 16 June, 2020, and proposing solutions to reduce the related risks". Most landslides are shallow, triggered by rainfall and anthropogenic activities such as road excavation. Figure 2 shows selected landslide photos of the study area collected from the field survey. In addition, non-landslide locations were selected from areas with no recorded history of landslide occurrences. Accordingly, 248 non-landslide sites were identified to complement the landslide inventory. Landslide and the non-landslide regions were demarcated by creating buffer zones around landslide areas, taking into account topographical and geo-environmental factors.

Figure 2. Photos of landslides identified from fieldwork (Photo source: Tran Van Phong)

3.1.2. Factors affecting landslides

Beyond the inventory of past and present landslides, conditioning factors are essential components in assessing landslide susceptibility. Given the geological and geo-environmental conditions of the study area, ten landslide affecting factors namely, geology, elevation (m), slope (degree), curvature, aspect, Topographic Wetness Index (TWI), river density (km/km^2), fault density (km/km^2), land cover, and Stream Power Index (SPI) were identified as contributors to landslide occurrence in this area. More specifically, elevation indirectly exerts a decisive influence on landslide occurrence by controlling related factors such as rainfall distribution, temperature variability, and both physical and chemical weathering (Yufeng and Fengxiang, 2009). Slope degree is one of the most important topographic parameters in landslide control; for this reason, it has been used in almost all landslide studies (García-Rodríguez and Malpica, 2010). Curvature shows the morphological and topographic curvature of the area (Šilhán, 2021). The effect of surface curvature on erosion due to water flow, and thus on the occurrence of landslides. Concave surfaces are more vulnerable to landslides due to water accumulation (Ercanoglu and Gokceoglu, 2002). The negative values indicate surface concavity, positive values indicate convexity, and zero values indicate surface smoothness (Ercanoglu and Gokceoglu, 2002). The direction of the slope is important for the sliding of rock/ground mass (Hong et al., 2017); thus, aspect is an important factor, as it affects the wetness of the ground due to the direction of sunlight, wind, and rainfall (Hong et al., 2017). Soil moisture and surface saturation are indicators of TWI, which can affect slope stability (Tien Bui et al., 2016; Wang et al., 2020). SPI indicates the stream's erosion capacity, which is directly related to slope and watershed area. Therefore, when surface flow velocity increases due to SPI-eroded terrain, the likelihood of mass movements increases. River density is one of

the controlling factors in slope stability (Schlögel et al., 2018). Faults are tectonic breaks that create sliding planes and form rock blocks. Direction, orientation, and nature of faults play essential roles in landslide activity. Thus, a fault density map was created, which is also considered one of the important landslide affecting factors (Chen et al., 2017b; Xu et al., 2012). Geology is one of the most important parameters in the study of landslides because different rock units exhibit varying degrees of susceptibility to landslides due to their mineralogy, weathering characteristics, and permeability (Xu et al., 2012; Zhuang et al., 2015; Prakash and Pham, 2023). In addition, land cover plays an essential role in slope stability, as bare ground is more prone to landslides than vegetation-covered areas (Abernethy and Rutherfurd, 2001).

In the present study, thematic maps of the influencing parameters were compiled from various sources. More specifically, topographical factors such as elevation (m), slope (degrees), curvature, aspect, TWI, and SPI were obtained by geoprocessing a Digital Elevation Model (DEM) with a spatial resolution of 12.5 m, extracted from the Alaska Satellite Facility database (https://asf.alaska.edu). Other factors, such as river density (km/km²) and fault density (km/km²), were derived from the map of major rivers (https://www.diva-gis.org/gdata). In contrast, the fault map was extracted from the geological map of the area (General Department of Geology of Vietnam). River density was prepared using the river layers extracted from the DEM (Schlögel et al., 2018). A geological map was obtained from the General Department of Geology of Vietnam, showing the geological groups (Table 1). The land cover map was collected from ESRI (https://livingatlas.arcgis.com/landcover/) (Fig. 3).

Table 1. Characteristics of the geological Groups of the study area

No	Symbols	Characteristics
1	G1	Quaternary: pebble, granule, sand, silty clay
2	G2	Dien Bien complex, phase 1: gabbro, gabbro
3	G3	Dien Bien complex, phase 3: granite, spessartite, porphyritic diorite
4	G4	Dien Bien complex, phase 3: biotite-hornblend granite
5	G5	Dien Bien complex, phase 2: granodiorite, diorite quartz
6	G6	Siltstone, silty sandstone, thin layers of coaly shale
7	G7	Dien Bien Complex, phase 3: granite; dyke phase: fine-grained granite, spessartite, and porphyritic diorite
8	G8	Lamprophyr
9	G9	Middle Song Da formation: tuff, basalt, andesite, and ryolite
10	G10	Lower Song Da formation: pebbles, grits, sandstones, quartzite siltstones, siliceous schist, shale, thin-layer or lenticular limestone, andesitobasalt, andesite, dacite, ryolite, felsite, and their tuffs
11	G11	Lower Subformation: conglomerate, gritstone, sandstone, siltstone, marl
12	G12	Phu Si Lung complex, phase 3: granite biotite, granite 2 mica
13	G13	Phu Si Lung complex, phase 2: small-grained leucocratic biotite granite, two-mica granite
14	G14	Phu Si Lung Complex, phase 1: melanocratic granite, porphyritic biotite granite, and compressed medium- to coarse-grained two-mica granite
15	G15	Granite biotite, granit 2 mica
16	G16	Phu Si Lung complex, phase 3: porphyritic biotite granite and two-mica granite
17	G17	Nam Man formation: Conglomerate, gritstone, sandstone, chocolate siltstone, and claystone
18	G18	Clay shale, sericitied clay shale, sandstone interbedded with quartzitic sandstone, marl, limestone
19	G19	Dien Bien complex, phase 3: biotite-hornblend granite
20	G20	Dien Bien complex, phase 1: granodiorite, diorite
21	G21	Quartz sandstone, clay shale, marl, a little conglomerate, and gritstone

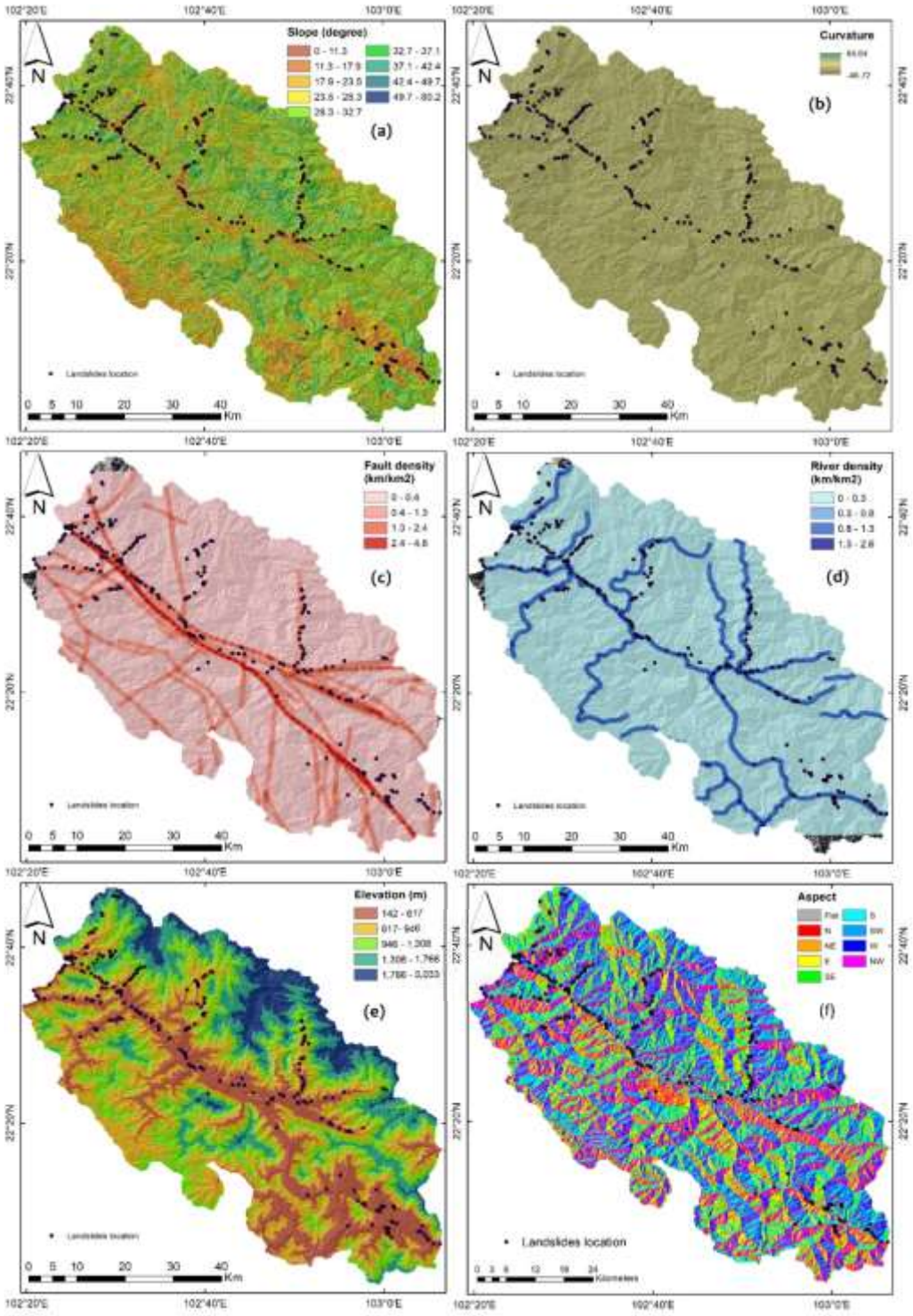

Figure 3. Conditioning factors used in this study: (a) slope, (b) curvature, (c) fault density, (d) river density, (e) elevation, (f) aspect, (g) SPI, (h) TWI, (i) land cover, and (j) geology

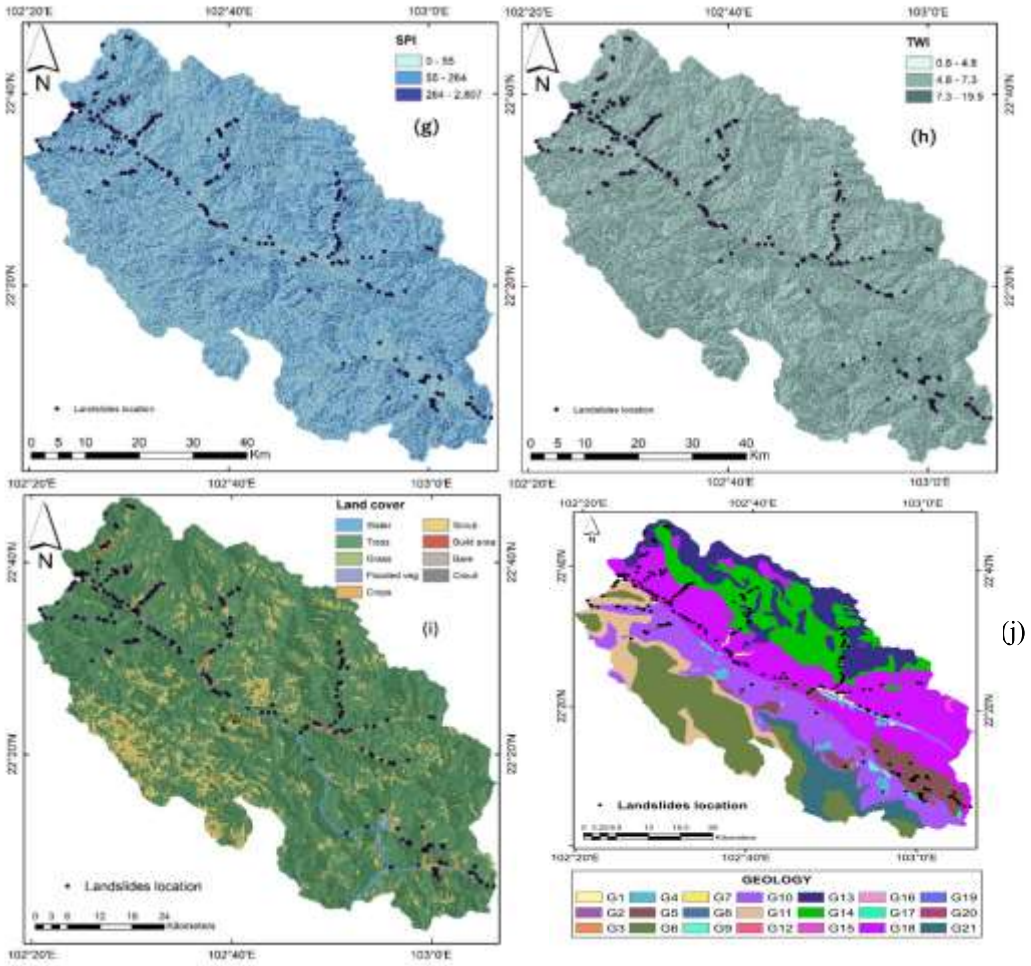

Figure 3. Cont.

3.2. Methods used

3.2.1. Partial Decision Trees (PART)

The decision tree algorithm is among the most widely used techniques in data mining. In data mining, a decision tree is an ML model that can be used for both regression and classification. It is one of the most widely used tools and techniques in data mining (Exarchos et al., 2012), especially when data volumes are very high (Dumitrescu et al., 2022). PART is a partial decision tree over all current samples, and the leaf with the highest coverage is selected for the new rule. Then, the partial decision tree is removed, preventing the initial generalization (Dauda et al., 2019). The steps for performing the algorithm are described in the work by Luu et al. (2021). The PART algorithm combines C4.5 and RIPPER classification models. It is worth noting that partial decision trees and association rules are effective and reliable decision-making techniques, providing a simple, interpretable representation of the extracted knowledge while achieving high classification accuracy.

In contrast to traditional decision tree models, the rule-based model generated by this kind of analysis can be easily interpreted by professionals without requiring specialized statistical knowledge. Additionally, like other conventional decision tree algorithms, the

PART model can analyze very high-dimensional feature spaces (Berger et al., 2006). PART was selected in this study in view of its advantage over other algorithms. In this algorithm, causal relationships between variables are analyzed using the direct relation matrix (Kareem and Jasim, 2022). This is one of the main reasons why the PART algorithm is preferred.

3.2.2. Decorate ensemble

Decorate is formed by a powerful learner with high precision on the training dataset. It creates a diverse and effective group in a simple framework (Nikmanesh et al., 2022). In this way, when creating new groups, different random samples are added to the training set. To these artificially created new specimens, a classification label is then assigned (Lei et al., 2019). The Decorate algorithm is particularly effective in several aspects: it reduces the number of required training samples while maintaining model accuracy, leverages unlabeled data to enhance performance within a semi-supervised learning framework, and integrates active and semi-supervised learning strategies to achieve improved predictive accuracy (Sopha et al., 2022). Unlabeled classification is performed in such a way that each class provides a base classifier (Ci) in the set of probabilities C for class x. If $P_{Ci;\ y}$ (x) is the probability that example x belongs to class Y with respect to the classifier Ci, the likelihood of being in each class for the whole set is calculated. In this work, Decorate was applied as an ensemble optimization method with PART as the base classifier for landslide susceptibility modeling.

3.2.3. Bagging Ensemble

Breiman (1996) has used the concept of Bootstrap Aggregating to create various estimates. This technique assumes that trained data is available for a community and that simulated scenarios are generated from this data. Therefore, resampling will provide the required diversity by using different datasets, and when a new sample is presented to each classifier, a majority vote is used to identify the class (Akila and Srinivasulu Reddy, 2018). Pasting is a Bagging method generally designed for large data sets. These datasets are split into smaller subsets for training the various classifiers. In this case, there are two types of random voting and importance voting: the first generates the first set of subsets randomly, and the second generates the second set in consecutive order based on the importance of these instances. Important examples are those that increase dataset diversity (Wu et al., 2018). The method relies on a weak and rigid data distribution to construct the dataset. Out-of-bag classifiers identify difficult instances; an instance is considered "hard" if it is misclassified by the ensemble (Yang et al., 2017a). These hard instances are consistently incorporated into subsequent datasets, whereas easier instances are less likely to be included (He et al., 2022). In this work, Bagging was applied as an ensemble optimization method, with PART serving as the core classifier, for landslide susceptibility modeling.

3.2.4. Logistic Regression (LR)

LR aims to show how quantitative or qualitative variables impact a two-dimensional (two-tier) dependent variable (Nibbering and Hastie, 2022). Unlike linear regression, which uses a small dependent variable, logistic regression uses a qualitative, two-dimensional dependent variable (Peng and Lu, 2021). In LR, qualitative independent variables must either be two-dimensional or treated as two-dimensional apparent variables (Ghazvini et al., 2019). In the LR method, a function called "Logistic Function" is used. For this reason, this regression method is called logistic regression (Donnelly and Verkuilen, 2017). In this study, Logistic

Regression (LR) was chosen as a benchmark single-machine learning model for comparison with hybrid models in landslide susceptibility mapping.

3.2.5. Validation indicators

The receiver operating characteristic (ROC) curve is plotted as a function of the actual positive rate and false positive rate (Schechter et al., 2010; Van Quang et al., 2025; Hai et al., 2022). The area under this curve, called the AUC, is often used for quantitative evaluation of classification models (Freyer et al., 2001; Prakash et al., 2024; Xuan et al., 2024). Values of AUC range from 0.5 to 1, with 0.5 indicating low performance and 1 indicating perfect performance.

SPE (Specificity) is the fraction of negative responses that are correctly detected, such as the percentage of landslides that do not exist as predicted by the model (Eckley and Tangerina, 2021). SST (Sensitivity) is the fraction of positive responses that are correctly detected, such as the percentage of landslides that occur according to the model prediction (Yin and Tian, 2014). ACC (Accuracy) means how close the measured value is to the actual value. High bias and variance mean low accuracy (Jablonski, 2020). PPV (Positive Predictive Value) is the ratio of correctly selected positive cases to the total number of positive cases (Lewandrowski, 2019). NPV (Negative Predictive Value) means the proportion of negative landslides that are correctly labeled as non-landslide (Duncan et al., 2020). RMSE (Root Mean Square Error) is a metric that quantifies the difference between the predicted and actual values (Schubert et al., 2017; Nguyen et al., 2023; Vu et al., 2021). The equations for the criteria are described in published works (Nhu et al., 2020; Phong et al., 2021). Model performance improves with increasing SPE, SST, ACC, PPV, and NPV, while lower RMSE indicates greater accuracy (Luu et al., 2022).

McNemar's test was also used to evaluate the statistical significance of the difference between the models in this study. It is based on the null hypothesis that the difference between the models is not statistically significant, corresponding to a p-value of 0.05 (Kavzoglu, 2017). In this case, the p-value is smaller than 0.05, the null-hypothesis is rejected, and vice versa. McNemar's test is used to determine the statistical significance of differences between classifiers (Lachenbruch, 2014).

3.2.6. OneR feature selection

OneR is regarded as a crucial step in landslide susceptibility modeling (Pham et al., 2021a). To accomplish this objective, subsets of the population are selected that are more likely to affect future predictions of events such as landslides. Known formally as "One Rule", OneR is an algorithm for creating precise classifications of data structures (Dung et al., 2021). It generates a rule for each anticipant in the data, then chooses the rule that produces the smallest error as its "Rule" for further classifications (Zia et al., 2015). A set of input variables is selected before the output variable is chosen based on the statistical relationship between them. The algorithm seeks a simple rule that applies to all factors by assigning each factor's value to the majority class (Gnana et al., 2016). Following the evaluation of rule accuracy, factors are assessed and ordered by their Average Merit (AM) index, which reflects the quality of the generated rules and indicates the quality of the corresponding rule.

4. Results and discussion

4.1. Evaluation of feature importance

In machine learning, feature selection techniques are employed to enhance algorithmic predictive power by retaining the most significant variables while discarding those that are non-informative. In the present study, the importance of landslide

conditioning factors was evaluated using the OneR feature selection method, as shown in Fig. 4. It can be observed that although all conditioning factors contribute to modeling landslide susceptibility. Still, the two factors, namely elevation (AM = 69.540) and river density (AM = 67.816), have the first and second highest importance among the factors. It is reasonable, as the elevation of an area can significantly influence the occurrence and frequency of landslides. Specifically, regions with moderate to steep slopes and high elevations are more susceptible to landslides, whereas areas with gentle slopes and lower elevations are less susceptible. This is because high elevations often have steeper slopes and are more prone to gravity- and water-driven erosion, which can destabilize soil and rock formations, leading to landslides.

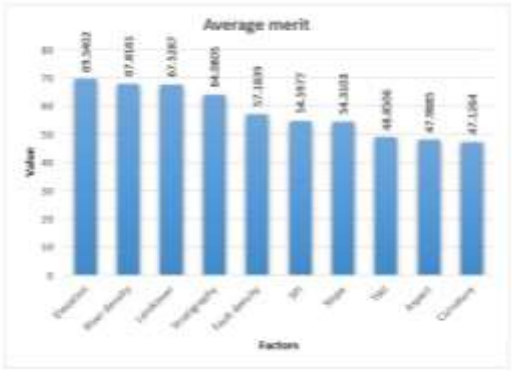

Figure 4. Feature importance of conditioning factors determined by OneR

Furthermore, river density plays a significant role in landslide susceptibility, with regions exhibiting high river density tending to experience more frequent landslides. This is because rivers can erode and destabilize slopes, particularly when the slopes are composed of loose or weak rock formations. The flow of water in a river can saturate the soil and rock, weakening their cohesion and increasing the risk of landslides. Moreover, soil moisture can reduce friction between soil particles, facilitating slope failure and downslope movement. Overall, the ten conditioning factors were carefully selected to predict landslide susceptibility in this study.

Here, we note that in the present study, we have not considered factors such as soil type, roads, NDVI, or rainfall. The reason is that roads are an essential feature, and most landslides occur along roads; thus, road excavation can be considered an anthropogenic triggering factor of instability in the rock mass. Similarly, rainfall acts as a natural trigger, as most landslides occur during the rainy season. We were unable to obtain a detailed soil-type map of the area, so we have not included it in this study. Moreover, we require a detailed soil map showing surface and subsurface geotechnical information for landslide studies, which is difficult to obtain at the regional scale. The NDVI factor has not been considered, as the study area is covered with dense vegetation. However, these parameters will be considered in future studies on availability.

4.2. Models' training and evaluation

The model's training process began with selecting the hyperparameters for each method. In this process, we used a trial-and-error approach to optimize and select the hyperparameters for each model. Table 2 shows the chosen hyperparameters for training the models.

Validation results demonstrate that all models achieved acceptable predictive performance; however, the hybrid models consistently outperformed the single models. In particular, the D-PART model showed the highest predictive capability on the validation dataset (AUC = 0.801), followed by B-PART (AUC = 0.795), PART (AUC = 0.758), and LR (AUC = 0.736). The superior performance of the ensemble-based models highlights the effectiveness of Decorate and Bagging techniques in reducing model variance and improving generalization ability (Figs. 5, 6, and 7).

Table 2. Hyperparameter settings of machine learning models used in this research

No	Hyperparameters	Model			
		LR	PART	D-PART	B-PART
1	Batch Size	100	100	100	100
2	The maximum number of iterations allowed	-1	-	-	-
3	Number of Decimal Places	4	2	2	2
4	Ridge	1.00E-08	-	-	-
5	Confidence Factor	-	0.25	-	-
6	Minimum number of instances per rule	-	2	-	-
7	Number of Folds	-	3		
8	Seed	-	1	1	1
9	Artificial Size	-	-	1	-
10	Desired Size	-	-	15	-
11	Number of Iterations	-	-	50	10
12	Bag Size Percent	-	-	-	100
13	Number of Execution Slots	-	-	-	1

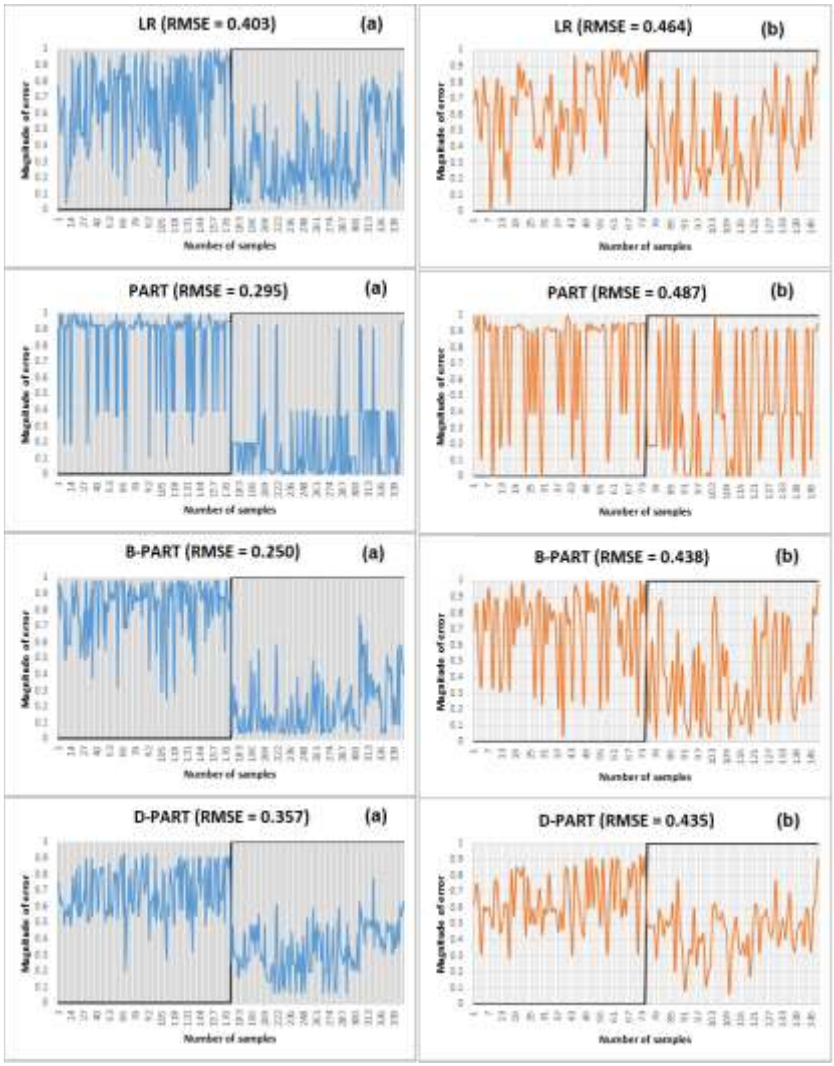

Figure 5. Model error analysis for (a) training and (b) validation datasets

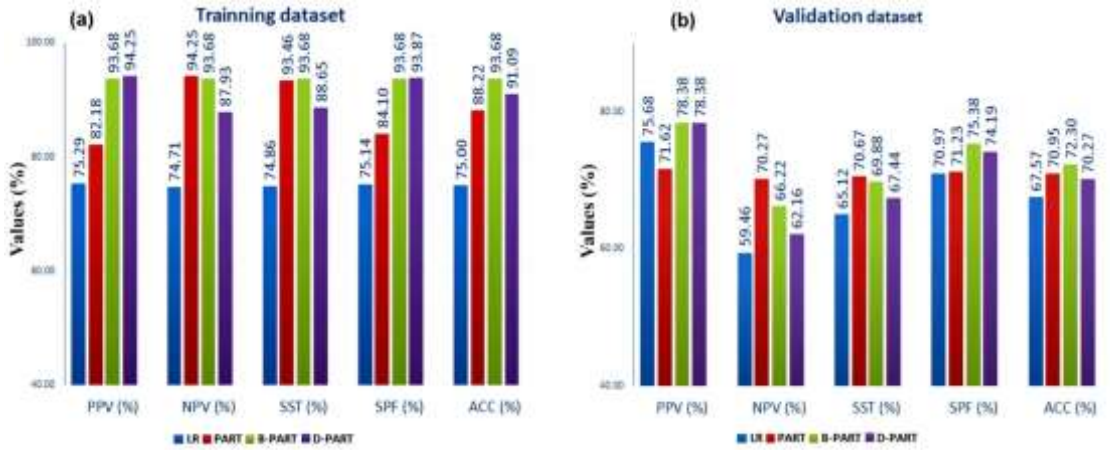

Figure 6. Evaluation of model performance using (a) training data and (b) validation data

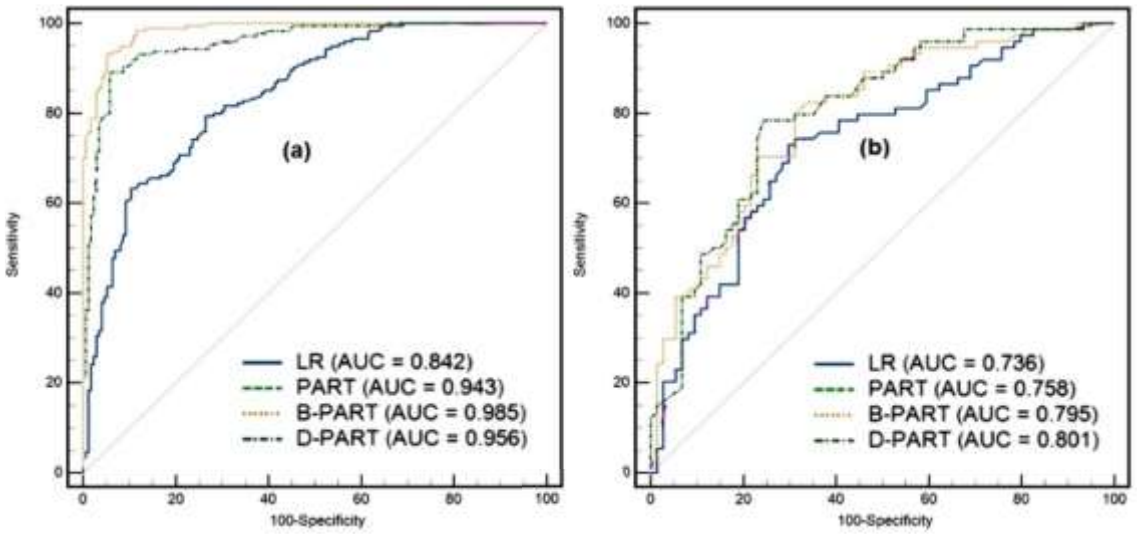

Figure 7. AUC-ROC of the models during (a) the training phase and (b) the validation phase

In general, the models developed and applied in this study exhibited satisfactory performance in predicting and mapping landslide susceptibility, but D-PART and B-PART is better compared with other models (PART and LR), and D-PART is slightly better than B-PART. It is reasonable, as both D-PART and B-PART used Decorate-and-Bagging ensembles to optimize the training dataset for classification. More specifically, the advantages of Decorate are (i) it can generate diversity by training multiple base models with different learning algorithms, initializations, and/or subsets of the training data; thus, it helps to reduce overfitting and improves generalization, (ii) it is more robust to noise and outliers in the data as the other models in the ensemble can compensate for the errors made by one model, (iii) it is easy to scale up to handle large

datasets and complex models as it is built by training multiple base models in parallel, which can be distributed across multiple machines. In the case of Bagging, it reduces variance by generating multiple bootstrap samples of the training data and training each base model on a different sample; thus, it helps reduce the effects of overfitting and improve the generalization performance of the predictor. In addition, it is more robust to changes in the training data or model parameters; thus, it is less likely to suffer from overfitting, underfitting, or unstable predictions. In this work, LR performed worst among the models, as one of its weaknesses is that it cannot solve nonlinear problems because the decision boundary is linear (Reed and Wu, 2013).

To further assess whether the observed differences in predictive performance among the models are statistically significant, McNemar's test was applied. The corresponding p-values for pairwise model comparisons on the training and validation datasets are presented in Tables 3 and 4, respectively. The results reveal that the performance differences between the hybrid ensemble models and the single models are statistically significant ($p < 0.05$) in most comparisons, particularly between D-PART and LR, and between D-PART and PART. Conversely, the differences between D-PART and B-PART are not statistically significant ($p > 0.05$), indicating that both ensemble-based models perform comparably. However, D-PART consistently achieves slightly higher predictive accuracy.

Table 3. P-values of McNemar's test of the models (training dataset)

Models	LR	PART	B-PART	D-PART
LR		0.0031	1	0.53
PART	0.0031		0.0002	0.0001
B-PART	1	0.0002		0.39
D-PART	0.53	0.0001	0.39	

Table 4. P-values of McNemar's test of the models (validation dataset)

Models	LR	PART	B-PART	D-PART
LR		0.0275	1	0.2478
PART	0.0275		0.0113	0.0001
B-PART	1	0.0113		0.1796
D-PART	0.2478	0.0001	0.1796	

4.3. Landslide susceptibility mapping

Using LR, PART, B-PART, and D-PART, landslide susceptibility maps were constructed (Fig. 9). As a first step, the susceptibility index was computed for every pixel across the study region during model validation. Thereafter, these indices were classified into five intervals using the Natural Breaks method, which is based on natural breaks in the data distribution (Mehrabi, 2021). Finally, landslide susceptibility maps were constructed with five classes (very high, high, moderate, low, and very low) corresponding to five classified intervals (Fig. 8). To validate the performance of the susceptibility maps, Frequency Ratio (FR) analysis was carried out on each susceptibility class of the maps. It is the ratio of the percentage of landslide or non-landslide pixels to the rate of susceptibility-class pixels. Results of the FR analysis are shown in Fig. 9. It can be seen that in the map generated by LR, the FR value is 2.09 in the very high class, followed by high (1.30), moderate (0.47), low (0.28), and very low (0.11), respectively. For the map generated by PART, the FR value is highest in the high

(3.15) and very high (3.14) categories, followed by moderate (1.57), low (0.42), and very low (0.20). According to the map generated by B-PART, the highest FR value is very high (3.77), followed by high (2.53), moderate (0.77), low (0.42), and very low (0.17). According to the map generated by D-PART, the highest FR value is very high (6.22), followed by high (3.71), moderate (1.31), low (0.50), and very low (0.11), respectively. In general, the frequency ratio (FR) is highest in areas classified as high and very high susceptibility across all maps, indicating that the landslide susceptibility maps generated by each model are reliable for practical susceptibility prediction.

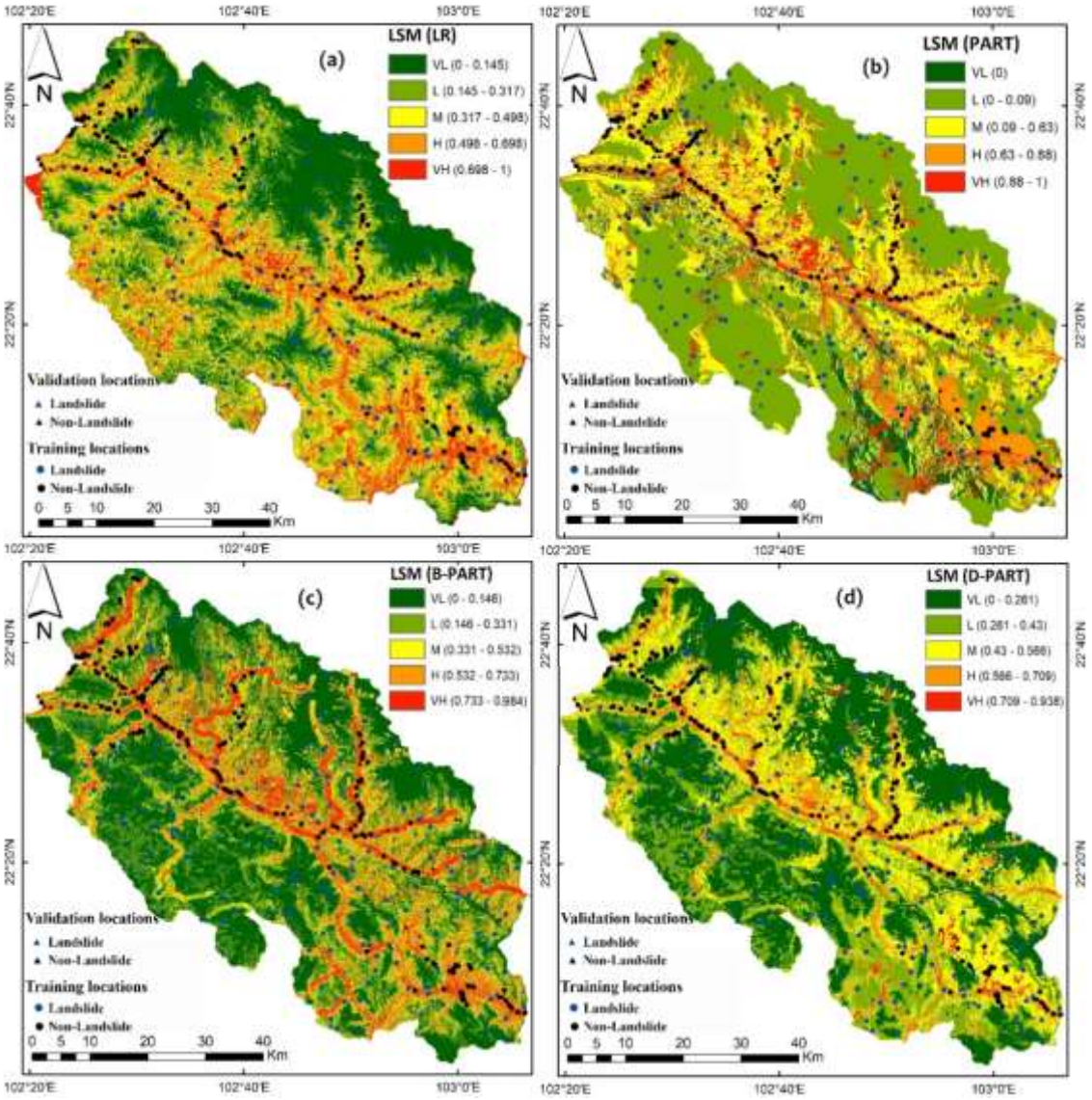

Figure 8. LSM built from various models: (a) LR, (b) PART, (c) B-PART, (d) D-PART

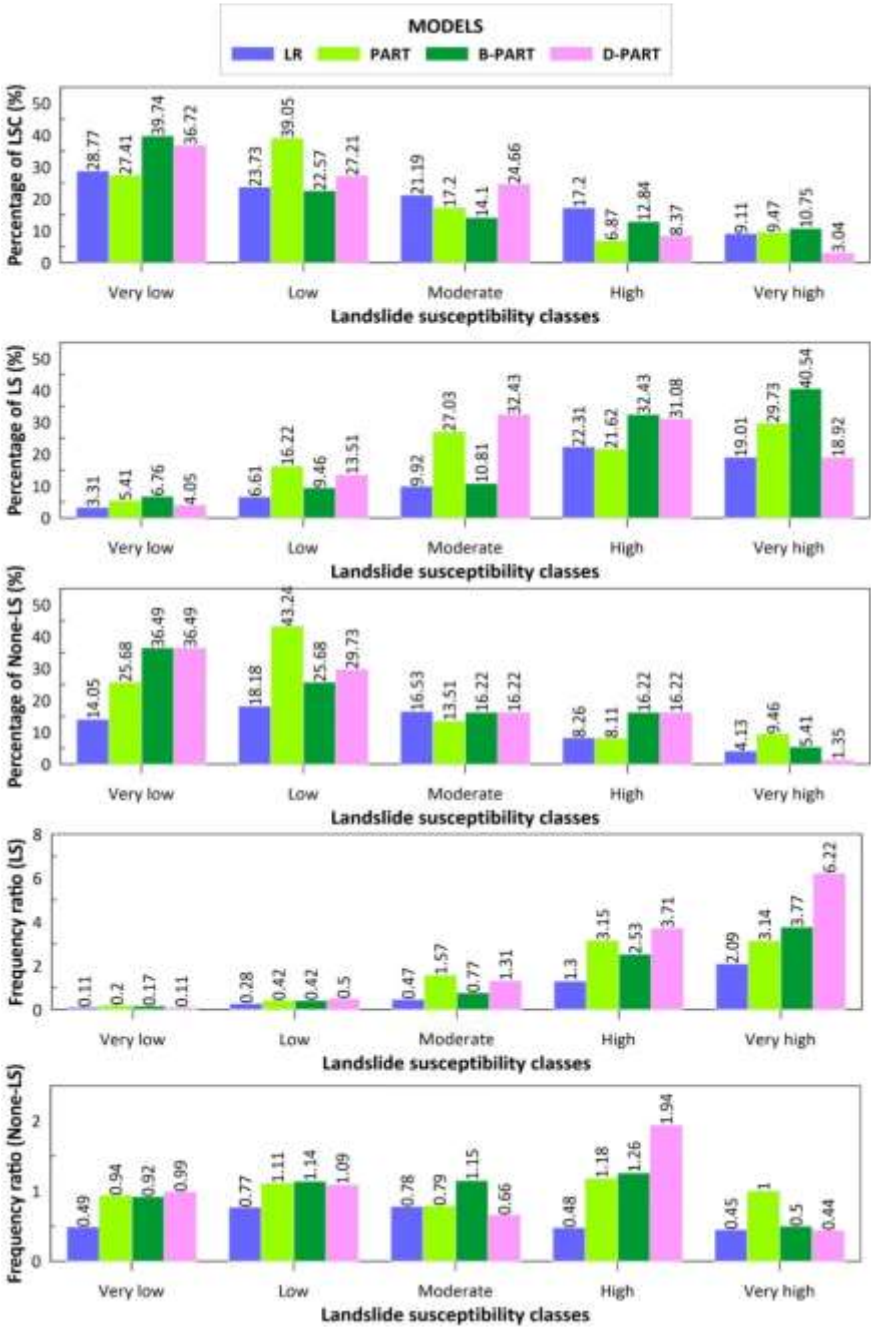

Figure 9. Frequency ratio (FR) analysis of the susceptibility maps
(LS: landslides and Non-LS: Non-landslides, LSC: Landslide susceptibility classes)

5. Concluding Remarks

Creating accurate landslide susceptibility maps of landslide-prone areas is a valuable task for landslide hazard and risk management, not only in hilly regions of Vietnam but also worldwide. There are several ML models available for this task, but

it is desirable to continue exploring the development of new hybrid models to accurately predict landslide-susceptible zones. With this objective, we have developed two novel models, D-PART and B-PART, for landslide susceptibility mapping in the Muong Te area, Lai Chau province, Vietnam, considering 10 landslide-affecting factors. OneR feature selection analysis indicated that all conditioning factors contribute to the modeling of landslide susceptibility. Still, the two features, namely elevation (AM = 69.5402) and river density (AM = 67.8161), are the most important factors in this area.

Although the proposed models demonstrate strong predictive performance, certain limitations should be acknowledged. Some potentially important triggering factors, such as rainfall intensity, detailed soil properties, road proximity, and vegetation indices (NDVI), could not be included due to data unavailability at the regional scale. Future studies may integrate these factors, apply the proposed hybrid models across different geomorphological settings, and explore additional ensemble or deep learning approaches to further enhance landslide susceptibility prediction.

Overall, this study confirms that ensemble-based hybrid machine learning models, particularly D-PART, provide improved predictive accuracy compared to conventional single models. The proposed methodology is transferable and can be effectively applied to other landslide-prone mountainous regions, provided appropriate local conditioning factors are selected.

Acknowledgments

This study was financially supported. by the Vietnam Academy of Science and Technology (VAST) under a grant Project codes CBCLCA.16/25-27 and CSCL12.03/25-26.

References

Abernethy B., Rutherfurd I.D., 2001. The distribution and strength of riparian tree roots in relation to riverbank reinforcement. Hydrological Processes, 15, 63. https://doi.org/10.1002/hyp.152.

Achour Y., Boumezbeur A., Hadji R., Chouabbi A., Cavaleiro V., Bendaoud E.A., 2017. Landslide susceptibility mapping using analytic hierarchy process and information value methods along a highway road section in Constantine, Algeria. Arabian Journal of Geosciences, 10, 1−16. https://doi.org/10.1007/s12517-017-2980-6.

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. https://doi.org/10.1016/j.gsf.2019.10.001.

Achour Y., Saidani Z., Touati R., Pham Q.B., Pal S.C., Mustafa F., Balik Sanli F., 2021. Assessing landslide susceptibility using a machine learning-based approach to achieving land degradation neutrality. Environmental Earth Sciences, 80, 1−20. https://doi.org/10.1007/s12665-021-09889-9.

Akila S., Srinivasulu Reddy U., 2018. Cost-sensitive Risk Induced Bayesian Inference Bagging (RIBIB) for credit card fraud detection. Journal of Computational Science, 27, 247−254. https://doi.org/10.1016/j.jocs.2018.06.009.

Althuwaynee O.F., Pradhan B., Park H.-J., Lee J.H.J.C., 2014. A novel ensemble bivariate statistical evidential belief function with knowledge-based analytical hierarchy process and multivariate statistical logistic regression for landslide susceptibility mapping. CATENA 114, 21−36. https://doi.org/10.1016/j.catena.2013.10.011.

Anbalagan R., 1992. Landslide hazard evaluation and zonation mapping in mountainous terrain. Engineering Geology, 32, 269−277. https://doi.org/10.1016/0013-7952(92)90053-2.

Anis Z., Wissem G., Vali V., Smida H., Essghaier G.M., 2019. GIS-based landslide susceptibility mapping using bivariate statistical methods in North-western Tunisia. Open Geosciences, 11(1), 708−726. Doi: 10.1515/geo-2019-0056.

Asadi Nalivan O., Mousavi Tayebi S.A., Mehrabi M., Ghasemieh H., Scaioni M., 2022. A hybrid intelligent model for spatial analysis of groundwater potential around Urmia Lake, Iran. Stochastic Environmental Research and Risk Assessment, 1−18. https://doi.org/10.1007/s00477-022-02368-y.

Berger H., Merkl D., Dittenbach M., 2006. Exploiting partial decision trees for feature subset selection in email categorization. In: Proceedings of the 2006 ACM symposium on Applied computing, 1105−1109. https://doi.org/10.1145/1141277.1141536.

Bordbar M., Paryani S., Pourghasemi H.R., 2022. Chapter 29 - Landslide spatial modeling using a bivariate statistical method in Kermanshah Province, Iran. In: Pourghasemi HR (ed) Computers in Earth and Environmental Sciences. Elsevier, 401−415. https://doi.org/10.1016/B978-0-323-89861-4.00026-9.

Chen W., Peng J., Hong H., Shahabi H., Pradhan B., Liu J., Zhu A.-X., Pei X., Duan Z., 2018. Landslide susceptibility modelling using GIS-based machine learning techniques for Chongren County, Jiangxi Province, China. Science of the Total Environment, 626, 1121−1135. https://doi.org/10.1016/j.scitotenv.2018.01.124.

Chen W., Pourghasemi H.R., Panahi M., Kornejady A., Wang J., Xie X., Cao S., 2017a. Spatial prediction of landslide susceptibility using an adaptive neuro-fuzzy inference system combined with frequency ratio, generalized additive model, and support vector machine techniques. Geomorphology, 297, 69−85. https://doi.org/10.1016/j.geomorph.2017.09.007.

Chen W., Pourghasemi H.R., Zhao Z., 2017b. A GIS-based comparative study of Dempster-Shafer, logistic regression and artificial neural network models for landslide susceptibility mapping. Geocarto International, 32(4), 367−385. Doi: 10.1080/10106049.2016.1140824.

Chen W., Xie X., Wang J., Pradhan B., Hong H., Bui D.T., Duan Z., Ma J., 2017c. A comparative study of logistic model tree, random forest, and classification and regression tree models for spatial prediction of landslide susceptibility. CATENA, 151, 147−160. https://doi.org/10.1016/j.catena.2016.11.032.

Dauda K.A., Pradhan B., Uma Shankar B., Mitra S., 2019. Decision tree for modeling survival data with competing risks. Biocybernetics and Biomedical Engineering, 39(3), 697−708. https://doi.org/10.1016/j.bbe.2019.05.001.

Donnelly S., Verkuilen J., 2017. Empirical logit analysis is not logistic regression. Journal of Memory and Language, 94, 28−42. https://doi.org/10.1016/j.jml.2016.10.005.

Dumitrescu E., Hué S., Hurlin C., Tokpavi S., 2022. Machine learning for credit scoring: Improving logistic regression with nonlinear decision-tree effects. European Journal of Operational Research, 297(3), 1178−1192. https://doi.org/10.1016/j.ejor.2021.06.053.

Duncan J.R., Purnell J.C., Stratton M.S., Pavlidakey P.G., Huang C., Phillips C.B., 2020. Negative predictive value of biopsy margins of dysplastic nevi: A single-institution retrospective review. Journal of the American Academy of Dermatology, 82(1), 87−93. https://doi.org/10.1016/j.jaad.2019.07.037.

Dung N.V., Hieu N., Phong T.V., Amiri M., Costache R., Al-Ansari N., Prakash I., Le H.V., Nguyen H.B.T., Pham B.T., 2021. Exploring novel hybrid soft computing models for landslide susceptibility mapping in Son La hydropower reservoir basin. Geomatics, Natural Hazards and Risk, 12(1), 1688−1714. https://doi.org/10.1080/19475705.2021.1943544.

Eckley C.A., Tangerina R., 2021. Sensitivity, Specificity, and Reproducibility of the Brazilian Portuguese Version of the Reflux Symptom Index. Journal of Voice, 35(1), 161.e115−161.e119. https://doi.org/10.1016/j.jvoice.2019.08.012.

Ercanoglu M., Gokceoglu C., 2002. Assessment of landslide susceptibility for a landslide-prone area (north of Yenice, NW Turkey) by fuzzy approach. Environmental Geology, 41(6), 720−730. Doi: 10.1007/s00254-001-0454-2.

Exarchos T.P., Tzallas A.T., Baga D., Chaloglou D., Fotiadis D.I., Tsouli S., Diakou M., Konitsiotis S., 2012. Using partial decision trees to predict Parkinson's symptoms: A new approach for diagnosis and therapy in patients suffering

from Parkinson's disease. Computers in Biology and Medicine, 42(2), 195−204. https://doi.org/10.1016/j.compbiomed.2011.11.008.

Feizizadeh B., Shadman Roodposhti M., Jankowski P., Blaschke T., 2014. A GIS-based extended fuzzy multi-criteria evaluation for landslide susceptibility mapping. Computers & Geosciences, 73, 208−221. https://doi.org/10.1016/j.cageo.2014.08.001.

Freyer G., Ligneau B., Tranchand B., Ardiet C., Souquet P.-J., Court-Fortune I., Riou R., Rebattu P., Morignat E., Boissel J.-P., Trillet-Lenoir V., Girard P., 2001. The prognostic value of etoposide area under the curve (AUC) at first chemotherapy cycle in small cell lung cancer patients: a multicenter study of the groupe Lyon-Saint-Etienne d'Oncologie Thoracique (GLOT). Lung Cancer, 31(2), 247−256. https://doi.org/10.1016/S0169-5002(00)00174-4.

García-Rodríguez M.J., Malpica J.A., 2010. Assessment of earthquake-triggered landslide susceptibility in El Salvador based on an Artificial Neural Network model. Nat Hazards Earth Syst Sci., 10(6), 1307−1315. Doi:10.5194/nhess-10-1307-2010.

Ghazvini K., Yousefi M., Firoozeh F., Mansouri S., 2019. Predictors of tuberculosis: Application of a logistic regression model. Gene Reports, 17, 100527. https://doi.org/10.1016/j.genrep.2019.100527.

Gnana D.A.A., Balamurugan S.A.A., Leavline E.J., 2016. Literature review on feature selection methods for high-dimensional data. International Journal of Computer Applications, 136(1), 9−17. Doi: 10.5120/ijca2016908317.

Hai H.D., Ngo H.T.T., Van P.T., Duc D.N., Avand M., Huu D.N., Amiri M., Van L.H., Prakash I., Thai P.B., 2022. Development and application of hybrid artificial intelligence models for groundwater potential mapping and assessment. Vietnam Journal of Earth Sciences, 410−429. https://doi.org/10.15625/2615-9783/17240

He N., Yongqiao W., Tao J., Zhaoyu C., 2022. Self-Adaptive bagging approach to credit rating. Technological Forecasting and Social Change, 175, 121371. https://doi.org/10.1016/j.techfore.2021.121371.

Hong H., Chen W., Xu C., Youssef A.M., Pradhan B., Tien Bui D., 2017. Rainfall-induced landslide susceptibility assessment at the Chongren area (China) using frequency ratio, certainty factor, and index of entropy. Geocarto International, 32(2), 139−154. Doi: 10.1080/10106049.2015.1130086.

Jablonski A., 2020. Improved algorithm for calculating high accuracy values of the Chandrasekhar function. Computer Physics Communications, 251, 107237. https://doi.org/10.1016/j.cpc.2020.107237.

Kareem M.I., Jasim M.N., 2022. DDOS attack detection using lightweight partial decision tree algorithm. In: 2022 International Conference on Computer Science and Software Engineering (CSASE). IEEE, 362−367. Doi: 10.1109/CSASE51777.2022.9759824.

Kavzoglu T., 2017. Object-oriented random forest for high resolution land cover mapping using quickbird-2 imagery. In: Handbook of neural computation. Elsevier, 607−619. https://doi.org/10.1016/B978-0-12-811318-9.00033-8.

Lachenbruch P.A., 2014. McNemar test. Wiley StatsRef: Statistics Reference Online. https://doi.org/10.1002/9781118445112.stat04876.

Lei S., Guo S., Sun X., Yu H., Xu F., Wan N., Wang Z., 2019. Capture and dissociation of dichloromethane on Fe, Ni, Pd and Pt decorated phosphorene. Applied Surface Science, 495, 143533. https://doi.org/10.1016/j.apsusc.2019.143533.

Lewandrowski K.-U., 2019. Retrospective analysis of accuracy and positive predictive value of preoperative lumbar MRI grading after successful outcome following outpatient endoscopic decompression for lumbar foraminal and lateral recess stenosis. Clinical Neurology and Neurosurgery, 179, 74−80. https://doi.org/10.1016/j.clineuro.2019.02.019.

Luu C., Nguyen D.D., Amiri M., Van P.T., Bui Q.D., Prakash I., Pham B.T., 2022. Flood susceptibility modeling using Radial Basis Function Classifier and Fisher's linear discriminant function. Vietnam Journal of Earth Sciences, 44(1), 55−72. https://doi.org/10.15625/2615-9783/16626.

Luu C., Pham B.T., Tran V.P., Costache R., Nguyen H.D., Amiri M., Bui Q.D., Nguyen L.T., Le V.H., Prakash I., Phan T.T., 2021. GIS-Based Ensemble Computational models for Flood Susceptibility Prediction in the

Quang Binh Province, Vietnam, 599, 126500. https://doi.org/10.1016/j.jhydrol.2021.126500.

Mehrabi M., 2021. Landslide susceptibility zonation using statistical and machine learning approaches in Northern Lecco, Italy. Natural Hazards, 1−37. https://doi.org/10.1007/s11069-021-05083-z.

Mehrabi M., Moayedi H., 2021. Landslide susceptibility mapping using artificial neural network tuned by metaheuristic algorithms. Environmental Earth Sciences, 80, 1−20. https://doi.org/10.1007/s12665-021-10098-7.

Nguyen D.D., Nguyen H.P., Vu D.Q., Prakash I., Pham B.T., 2023. Using GA-ANFIS machine learning model for forecasting the load bearing capacity of driven piles. Journal of Science and Transport Technology, 3(2), 26−33. https://doi.org/10.58845/jstt.utt.2023.en.3.2.26-33.

Nhu V.-H., Mohammadi A., Shahabi H., Ahmad B.B., Al-Ansari N., Shirzadi A., Clague J.J., Jaafari A., Chen W., Nguyen H., 2020. Landslide Susceptibility Mapping Using Machine Learning Algorithms and Remote Sensing Data in a Tropical Environment. International Journal of Environmental Research and Public Health, 17(14), 4933. Doi: 10.3390/ijerph17144933.

Nibbering D., Hastie T.J., 2022. Multiclass-penalized logistic regression. Computational Statistics & Data Analysis, 169, 107414. https://doi.org/10.1016/j.csda.2021.107414.

Nikmanesh S., Safaiee R., Sheikhi M.H., 2022. Enhanced methane-sensing performances of Pd (monomer or dimer) decorated or doped γ-Graphyne: A DFT insight. Surfaces and Interfaces, 29, 101658. https://doi.org/10.1016/j.surfin.2021.101658.

Nohani E., Moharrami M., Sharafi S., Khosravi K., Pradhan B., Pham B.T., Lee S., M. Melesse A., 2019. Landslide Susceptibility Mapping Using Different GIS-Based Bivariate Models. Water, 11(7), 1402. https://doi.org/10.3390/w11071402.

Peng R., Lu Z., 2021. Semiparametric Averaging of Nonlinear Marginal Logistic Regressions and Forecasting for Time Series Classification. Econometrics and Statistics, 31, 19−37. https://doi.org/10.1016/j.ecosta.2021.11.001.

Pham B.T., Van D.D., Acharya T.D., Van P.T., Costache R., Van L.H, Nguyen H.B.T., Prakash I., 2021a. Performance assessment of artificial neural network using chi-square and backward elimination feature selection methods for landslide susceptibility analysis. Environmental Earth Sciences, 80, 1−13. https://doi.org/10.1007/s12665-021-09998-5.

Pham Q.B., Achour Y., Ali S.A., Parvin F., Vojtek M., Vojteková J., Al-Ansari N., Achu A., Costache R., Khedher K.M., 2021b. A comparison among fuzzy multi-criteria decision making, bivariate, multivariate and machine learning models in landslide susceptibility mapping. Geomatics, Natural Hazards and Risk, 12(1), 1741−1777. https://doi.org/10.1080/19475705.2021.1944330.

Phong T.V., Phan T.T., Prakash I., Singh S.K., Shirzadi A., Chapi K., Ly H.-B., Ho L.S., Quoc N.K., Pham B.T., 2021. Landslide susceptibility modeling using different artificial intelligence methods: a case study at Muong Lay district, Vietnam. Geocarto International, 36(15), 1685−1708. Doi: 10.1080/10106049.2019.1665715.

Prakash I., Nguyen D.D., Tuan N.T., Van P.T, Van H.L., 2024. Landslide susceptibility zoning: integrating multiple Intelligent models with SHAP Analysis. Journal of Science and Transport Technology, 23−41. https://doi.org/10.58845/jstt.utt.2024.en.4.1.23-41.

Prakash I., Pham B.T., 2023. Geotechnical Evaluation of Basalt Rocks: a review in the context of the construction of Civil Engineering structures. Journal of Science and Transport Technology, 10−24. https://doi.org/10.58845/jstt.utt.2023.en.3.4.10-24.

Prakash I., Pham B.T., 2024. Geotechnical Forensic Investigations of a Gravity Dam: Addressing Seepage and Sliding Problems in the Basalt Foundations of Karjan Dam, Gujarat, India. Journal of Science and Transport Technology, 24−38. https://doi.org/10.58845/jstt.utt.2024.en.4.3.24-38.

Reed P., Wu Y., 2013. Logistic regression for risk factor modelling in stuttering research. Journal of Fluency Disorders, 38(2), 88−101. https://doi.org/10.1016/j.jfludis.2012.09.003.

Sameen M.I., Sarkar R., Pradhan B., Drukpa D., Alamri A.M., Park H.-J., 2020. Landslide

spatial modelling using unsupervised factor optimisation and regularised greedy forests. Computers & Geosciences, 134, 104336. https://doi.org/10.1016/j.cageo.2019.104336.

Schechter T., Teuffel O., Gibson P., Sung L., Seto W., Gassas A., Doyle J., Dupuis L.L., 2010. Cyclosporine Area Under The Curve (AUC) In Children Undergoing Haematopoietic Stem Cell Transplantation (HSCT): Limited Sampling Strategy (LSS). Biology of Blood and Marrow Transplantation, 16(2), S258−S259. https://doi.org/10.1016/j.bbmt.2009.12.315.

Schlögel R., Marchesini I., Alvioli M., Reichenbach P., Rossi M., Malet J.P., 2018. Optimizing landslide susceptibility zonation: Effects of DEM spatial resolution and slope unit delineation on logistic regression models. Geomorphology, 301, 10−20. https://doi.org/10.1016/j.geomorph.2017.10.018.

Schubert A.-L., Hagemann D., Voss A., Bergmann K., 2017. Evaluating the model fit of diffusion models with the root mean square error of approximation. Journal of Mathematical Psychology, 77, 29−45. https://doi.org/10.1016/j.jmp.2016.08.004.

Shahabi H., Hashim M., 2015. Landslide susceptibility mapping using GIS-based statistical models and Remote sensing data in tropical environment. Scientific Reports, 5(1), 9899. Doi: 10.1038/srep09899.

Šilhán K., 2021. Spatial dendrogeomorphic sampling based on the specific tree growth responses induced by the landslide mechanism. Quaternary Geochronology, 61, 101132. https://doi.org/10.1016/j.quageo.2020.101132.

Sopha H., Ghigo C., Ng S., Alijani M., Hromadko L., Michalicka J., Djenizian T., Macak J.M., 2022. TiO_2 nanotube layers decorated by titania nanoparticles as anodes for Li-ion microbatteries. Materials Chemistry and Physics, 276, 125337. https://doi.org/10.1016/j.matchemphys.2021.125337.

Tien Bui D., Tuan T.A., Hoang N.-D., Thanh N.Q.C., Nguyen D.B., Liem N.V., Pradhan B., 2016. Spatial prediction of rainfall-induced landslides for the Lao Cai area (Vietnam) using a hybrid intelligent approach of least squares support vector machines inference model and artificial bee colony optimization. Landslides, 14, 447−458. https://doi.org/10.1007/s10346-016-0711-9.

Van Q.V., Van T.N., Dang B.H., Thang V.T., 2025. Real-time Vehicle Behavior Classification using Single 3-Axis Magnetic Field Sensor and Neural Network. Journal of Science and Transport Technology, 98−111. https://doi.org/10.58845/jstt.utt.2025.en.5.3.98-111.

Vu D.Q., Nguyen D.D., Bui Q.-A.T., Trong D.K., Prakash I., Pham B.T., 2021. Estimation of California bearing ratio of soils using random forest based machine learning. Journal of Science and Transport Technology, 45−58. https://doi.org/10.58845/jstt.utt.2021.en.1.1.45-58.

Wang G., Chen X., Chen W., 2020. Spatial Prediction of Landslide Susceptibility Based on GIS and Discriminant Functions. ISPRS International Journal of Geo-Information, 9(3), 144. https://doi.org/10.3390/ijgi9030144.

Wu Z., Li N., Peng J., Cui H., Liu P., Li H., Li X., 2018. Using an ensemble machine learning methodology-Bagging to predict occupants' thermal comfort in buildings. Energy and Buildings, 173, 117−127. https://doi.org/10.1016/j.enbuild.2018.05.031.

Xu C., Xu X., Dai F., Xiao J., Tan X., Yuan R., 2012. Landslide hazard mapping using GIS and weight of evidence model in Qingshui River watershed of 2008 Wenchuan earthquake struck region. Journal of Earth Science, 23(1), 97−120. Doi: 10.1007/s12583-012-0236-7.

Xuan B.T., Thuy D.L., Van P.T., Hong N.V., Van L.H., Nguyen D.D., Prakash I., Thanh T.P., Thai B.P., 2024. Groundwater potential zoning using Logistics Model Trees based novel ensemble machine learning model. Vietnam Journal of Earth Sciences, 46(2), 272−281. https://doi.org/10.15625/2615-9783/20316.

Yang K., Tian F., Chen L., Li S., 2017a. Realized volatility forecast of agricultural futures using the HAR models with bagging and combination approaches. International Review of Economics & Finance, 49, 276−291. https://doi.org/10.1016/j.iref.2017.01.030.

Yang W., Zheng Z., Zhang X., Tan B., Li L., 2017b. Analysis of landslide risk based on fuzzy extension

analytic hierarchy process. Journal of Intelligent & Fuzzy Systems, 33, 2523−2531. Doi: 10.3233/JIFS-17740.

Yin J., Tian L., 2014 Joint inference about sensitivity and specificity at the optimal cut-off point associated with Youden index. Computational Statistics & Data Analysis, 77, 1−13. https://doi.org/10.1016/j.csda.2014.01.021.

Yufeng S., Fengxiang J., 2009. Landslide Stability Analysis Based on Generalized Information Entropy. Paper presented at the Proceedings of the 2009 International Conference on Environmental Science and Information Application Technology, 2, 83−85. Doi: 10.1109/esiat.2009.258.

Zhuang J., Peng J., Iqbal J., Liu T., Liu N., Li Y., Ma P., 2015. Identification of landslide spatial distribution and susceptibility assessment in relation to topography in the Xi'an Region, Shaanxi Province, China. Frontiers of Earth Science, 9(3), 449−462. Doi: 10.1007/s11707-014-0474-3.

Zia T., Akhter M.P., Abbas Q., 2015. Comparative study of feature selection approaches for Urdu text categorization. Malaysian Journal of Computer Science, 28(2), 93−109.