Acta Geodyn. Geomater., Vol. 23, No. 1 (221), 129–149, 2026
DOI: 10.13168/AGG.2026.0008
journal homepage: https://www.irsm.cas.cz/acta

ORIGINAL PAPER

LANDSLIDE SUSCEPTIBILITY MAPPING OF THE HA LONG – VAN DON HIGHWAY USING NOVEL ENSEMBLE MODELS BASED ON DUAL PERTURB AND COMBINE FOR TREE-BASED (DPCT)

Tuan-Nghia DO [1) *], Tran Van PHONG [2, 3)], Phan Trong TRINH [2, 3)], Bui Nhi THANH [2)] and Vuong Hong NHAT[2)]

[1)] *Faculty of Civil Engineering, Thuyloi University, Hanoi, Vietnam*
[2)] *Institute of Earth Sciences, Vietnam Academy of Science and Technology, 84 Chua Lang Street, Hanoi, Vietnam*
[3)] *Graduate University of Science and Technology, Vietnam Academy of Science and Technology, 18 Hoang Quoc Viet Street, Hanoi, Vietnam*

**Corresponding author's e-mail:* dotuannghia@tlu.edu.vn

ARTICLE INFO

Article history:
Received 30 October 2025
Accepted 24 January 2026
Available online 4 February 2026

Keywords:
Landslide;
Machine Learning;
Ensemble model;
DPCT;
Ha Long – Van Don highway

ABSTRACT

The areas along transportation routes constructed in mountainous terrain often harbor significant landslide hazards. Ensemble learning techniques have proven their effectiveness in improving landslide susceptibility prediction performance. In this study, novel ensemble models (Bagging (B), Cascade Generalization (CG), and Dagging (D)) based on the Dual Perturb and Combine for Tree-based (DPCT) approach were employed to predict landslide susceptibility along the Ha Long – Van Don highway. The dataset comprised 77 landslide locations (3263 points), non-landslide locations (1:1 ratio with landslide points), and 14 conditional factors, including topography characteristics, geology, rainfall, and land use/land cover (LULC), which were input parameters for the models (B-DPCT, CG-DPCT, D-DPCT, and DPCT). Evaluation criteria for model prediction outcomes included the area under the receiver operating characteristic curve (AUC), parameters derived from the confusion matrix, the Kappa statistic, and the root mean square error (RMSE). The results demonstrate that the integration of higher-resolution datasets with hybrid machine-learning models leads to a significant improvement in predictive performance and accuracy for landslide susceptibility mapping compared to previous studies. Accordingly, landslide susceptibility maps predicted based on the B-DPCT model exhibited optimal evaluation results on the validation dataset (AUC = 0.948, accuracy ACC = 83.6, Kappa statistic = 0.67, and RMSE = 0.37), suggesting their recommended use for construction planning and mitigation efforts along the Ha Long – Van Don highway to minimize landslide-induced damages.

1. INTRODUCTION

In recent years, climate change and global warming have intensified extreme weather events, increasing both their frequency and magnitude (Ado et al., 2022; Azarafza et al., 2021; D'Amato and Akdis, 2020; Ghiasi et al., 2023; Ogunbode et al., 2020; Ullah et al., 2025). This escalation has led to a corresponding rise in the frequency and scale of natural disasters (AghaKouchak et al., 2020; Farinós-Dasí et al., 2024; Masson-Delmotte et al., 2022; Tin et al., 2024; Ward et al., 2020). Over the past two decades, natural disasters have increased by approximately 75 % globally, resulting in over 1 million fatalities, affecting more than 4 billion people, and causing economic losses nearing 3 trillion USD (UNDRR, 2022). Among these disasters, landslides significantly impact socio-economic systems (Yadav et al., 2023). Therefore, research on landslide prediction remains critical for disaster prevention and mitigation (He et al., 2025; Le Minh et al., 2023; Pham et al., 2022; Tong et al., 2023; Turker, 2025).

Landslide susceptibility mapping (LSM) is an effective tool in landslide forecasting, aiding governments in land management, urban planning, and settlement strategies (Ado et al., 2022; Ghiasi et al., 2023; He et al., 2025; Luat et al., 2024; Ullah et al., 2025; Singh et al., 2025; Li et al., 2025; Demirel et al., 2025). LSM employs two primary approaches: qualitative and quantitative. Quantitative methods, particularly those leveraging machine learning (ML), have proven more effective than qualitative methods (Saha et al., 2025; He et al., 2025; Ullah et al., 2025; Saha et al., 2023; Asadi et al., 2022; Ibrahim et al., 2020; Shano et al., 2020).

Ensemble ML techniques enhance the performance of single models in LSM applications (Li et al., 2025; Cao et al., 2025; Saha et al., 2024; Dey et al., 2024; Liu et al., 2023; Saha et al., 2022; Luat et al., 2020; Phi et al., 2024; Samadi et al., 2024). Commonly used ensemble techniques include Bagging (Gu et al., 2024; Zhang et al., 2024), Cascade Generalization (Ali et al., 2024; Hong, 2023a), Dagging (Bui et al., 2023; Le Minh et al., 2023; Tong et al., 2023), Decorate (Hong, 2023b; Le Minh et al., 2023), Multi Boost (Ajin et al., 2022; Bien et al., 2023) and Rotation Forest (Ali et al., 2024; Fang et al., 2021; Kalantar et al., 2020; Pham et al., 2022). Given the complexity and variability of landslide prediction across different

Cite this article as: Do T-N, Phong TV, Trinh PT, Thanh BN, Nhat VH: Landslide susceptibility mapping of the Ha Long – Van Don Highway using novel ensemble models based on dual perturb and combine for Tree-based (DPCT). Acta Geodyn. Geomater., 23, No. 1 (221), 129–149, 2026. DOI: 10.13168/AGG.2026.0008

regions, no single technique universally excels (Liu et al., 2023) (Costanzo and Irigaray, 2020; Ramos-Bernal et al., 2021). This variability underscores the need to explore and optimize new ML models for LSM applications (Ali et al., 2024). Each machine-learning model has its theoretical foundation and techniques (Azevedo et al., 2024). Therefore, tuning the parameters of these models helps improve accuracy in landslide prediction (Yu et al., 2024). Particularly, ensemble techniques prove effective in enhancing the prediction accuracy of the original single model (Tang et al., 2023; Zeng et al., 2023).

Highways are critical infrastructure, supporting socio-economic activities by enabling efficient transportation (He et al., 2025; Ngewie, 2024; Ullah et al., 2025; Zhou et al., 2024). Landslides along highways disrupt traffic, endanger lives, and damage infrastructure (Sun et al., 2023). Highway construction in mountainous areas often disturbs slopes, increasing landslide risks (Nguyen et al., 2020; Pasang and Kubíček, 2020). Furthermore, inadequate consideration in designing landslide mitigation structures contributes to frequent occurrences of landslides (Nguyen, Tien and Do, 2020; Van Tien et al., 2021). Predicting landslides along highway corridors is thus essential for risk reduction and infrastructure protection (Pasang and Kubíček, 2020; Sassa et al., 2020). While highway-corridor LSM has been studied globally, there remains a critical need for high-resolution, automated frameworks that can address the specific geotechnical instabilities of newly engineered cut-slopes in tropical regions.

The Halong - Vandon Highway is a newly constructed road that has been in operation only since February 2019; thus, landslides still occur frequently and remain a potential hazard in the future (Nguyen et al., 2020). Nguyen et al. (2020) and Van Tien et al. (2021) conducted preliminary assessments of the causes of landslides in the study area, identifying intense rainfall and steep cut slopes from road construction as the primary factors (Nguyen et al., 2020; Van Tien et al., 2021). More recently, Luat et al. (2024) carried out a study to produce a landslide susceptibility map along the Halong - Vandon Highway using an integration of GIS and the Analytic Hierarchy Process (AHP) and Frequency Ratio (FR) methods (Luat et al., 2024). However, previous works in this corridor often relied on qualitative weighting or general ML models that do not fully capture the local terrain details (Wadadar and Mukhopadhyay, 2022; Widiastuti et al., 2025).

The novelty and significant contributions of this study are summarized as follows. We propose a novel Dual Perturb and Combine for Tree-based (DPCT) ensemble framework that simultaneously incorporates data perturbation through bootstrap resampling and model perturbation using heterogeneous tree-based learners. This combination has not been jointly implemented in previous landslide susceptibility studies. In contrast to conventional ensemble techniques such as Bagging, Boosting, and Stacking, the DPCT framework enhances model diversity more effectively, mitigates overfitting under complex mountainous and anthropogenically modified terrain conditions, and improves model generalization capability. The proposed DPCT-based ensemble models were systematically evaluated and compared with individual machine learning models as well as traditional ensemble approaches, demonstrating consistent and significant improvements in predictive performance in terms of AUC, accuracy, and model stability. Moreover, this study represents one of the first applications of DPCT-based ensemble learning for landslide susceptibility mapping along a critical transportation corridor in Vietnam, providing meaningful scientific and practical implications for highway infrastructure planning and landslide risk management.

2. STUDY AREA

The Ha Long – Van Don Highway, constructed between September 2015 and December 2018, spans 59 km in Quang Ninh Province, northeastern Vietnam. This four-lane highway, designed for speeds of 100 km/h, connects key tourist destinations, including Ha Long Bay and Van Don Island, fostering regional socio-economic development (https://www.quangninh.gov.vn/). The study area spans 180.55 km² along the highway (Fig. 1), characterized by diverse topography with elevations ranging from 2.5 to 395 m. The climate includes a prolonged rainy season from May to October, with an average annual rainfall of 2,300 mm, a temperature of 23 °C, and humidity of 84.6 % (Technology, 2009). Geologically, the region comprises formations such as the Hon Gai Formation, Tan Mai Formation, Binh Lieu Formation, Ha Coi Formation, Cat Ba Formation, and the Quaternary (Fig. 4c), rich in coal seams, which may influence slope stability (Thanh, 2011). Since its operation, the highway has experienced recurrent landslides, posing risks to traffic safety (Fig. 2) (Van Tien et al., 2021).

3. METHODOLOGY AND MATERIALS

3.1. METHODOLOGY

The methodology for establishing LSM in the area of the Ha Long - Van Don highway is outlined in Figure 3. This study utilized machine learning models, including Dual perturb and combine for tree-based (DPCT), Bagging, Cascade Generalization (CG), Dagging, implemented in Weka 3.8.6 (https://ml.cms.waikato.ac.nz//weka/), with hyperparameters detailed in in Table 1.

3.2. ADOPTED MODELS

DPCT

The Dual Perturb and Combine for Tree-based (DPCT) algorithm, introduced by (Geurts and Wehenkel, 2005), is an advanced machine learning technique designed for classification and regression tasks. It builds upon the traditional Dual Perturb and

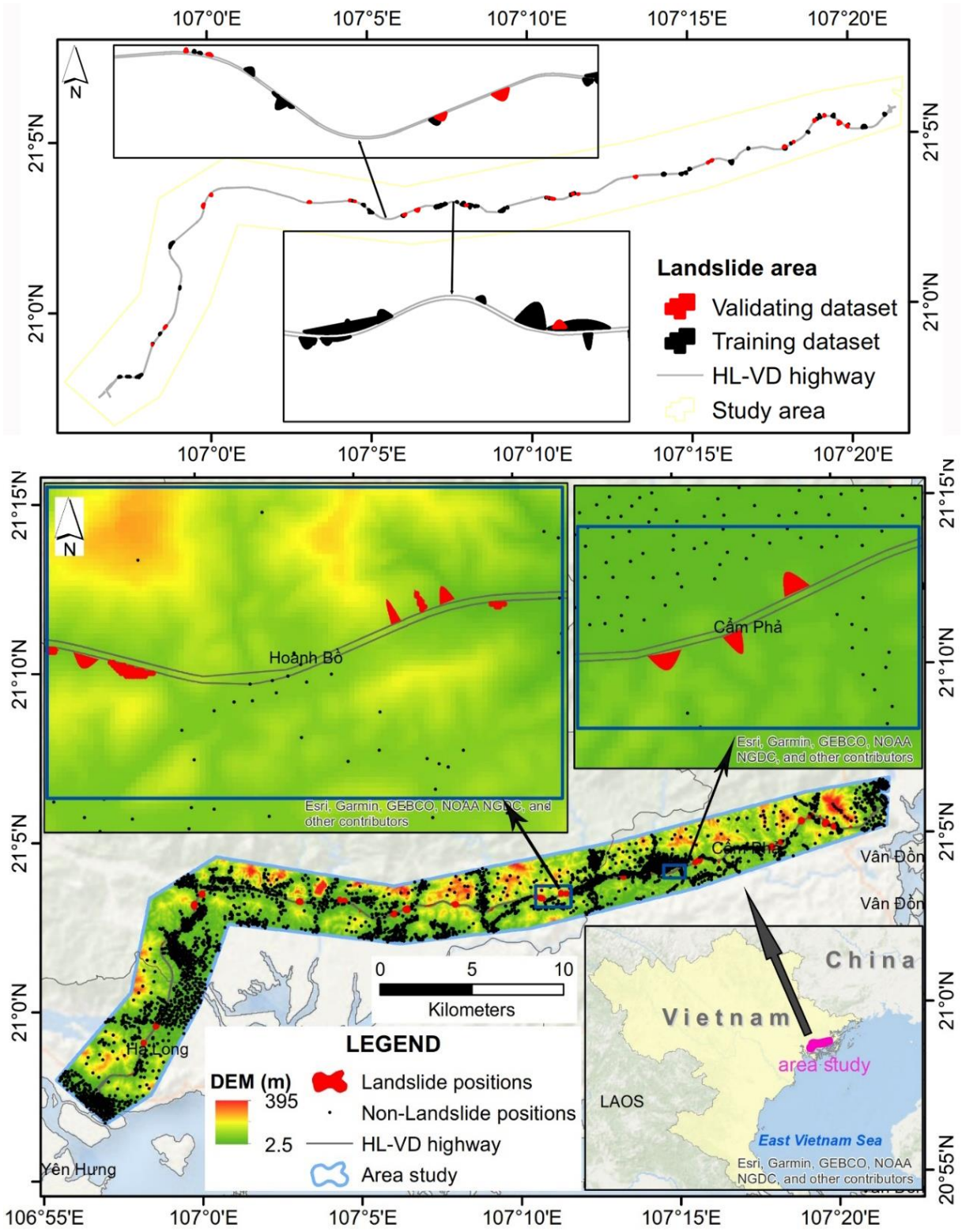

Fig. 1 Study area of Ha Long – Van Don Highway for landslide susceptibility mapping.

Fig. 2 Typical landslides along the Ha Long - Van Don highway: a) at km 10, b) at km 19, c) at km 24, and d) at km 30 (photo source: Tuan-Nghia Do).

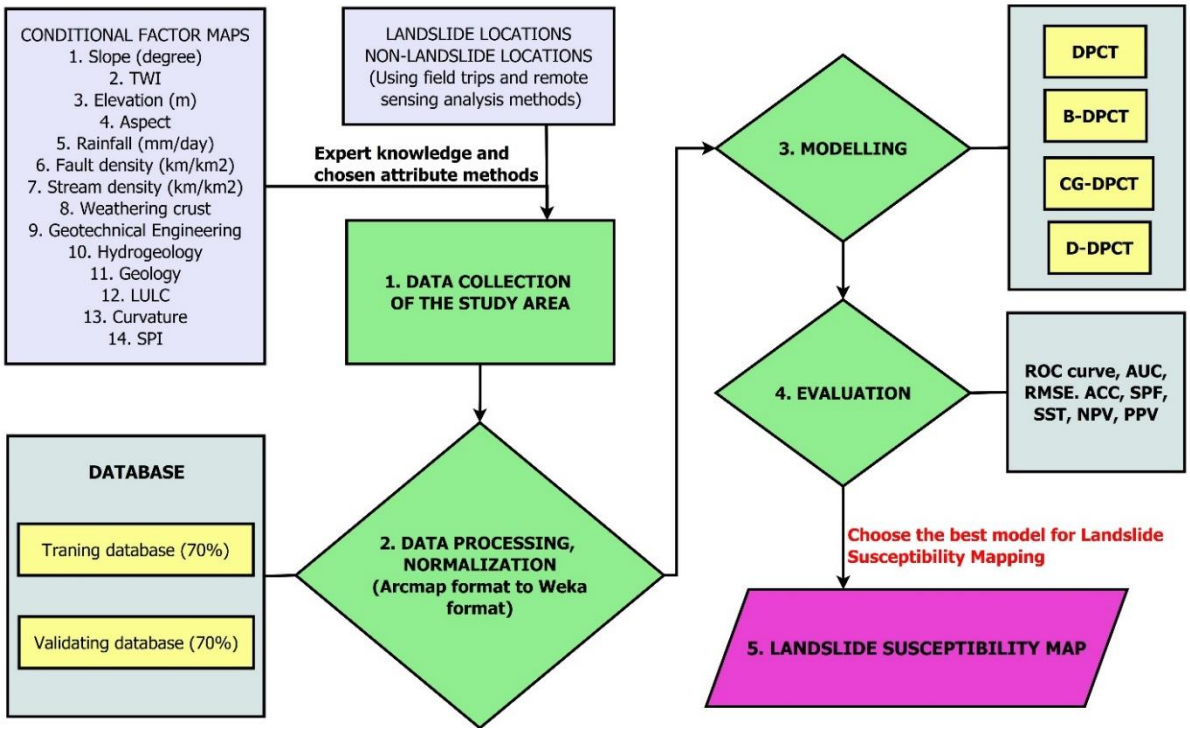

Fig. 3 The methodology for landslide susceptibility mapping along Ha Long – Van Don Highway.

Table 1 The parameters used for establishing landslide susceptibility maps in the study area.

No	Hyperparameters	Models			
		DPCT	**Bagging**	**Cascade Generalization**	**Dagging**
1	Lambda	0.2	-	-	-
2	Batch Size	100	100	100	100
3	Classifier	-	DPCT	DPCT	DPCT
4	Number of Decimal Places	-	2	2	2
5	Number of Execution Slots	-	1	1	-
6	Number of Interations	-	10	-	-
7	Seed	-	1	1	1
8	Number of Folds	-	-	20	2

Combine (DPC) framework (Geurts, 2001; Geurts and Wehenkel, 2005) by analytically combining predictions from perturbed datasets, eliminating the need for iterative training and prediction cycles. This analytical approach optimizes combination weights to enhance efficiency and compatibility with tree-based models (Geurts and Wehenkel, 2005; Khosravi et al., 2022). The operational sequence of DPCT is as follows:

Step 1: Data Perturbation: Generate multiple perturbed versions of the original dataset using techniques such as bootstrapping, feature sampling, or instance sampling to introduce controlled randomness.

Step 2: Base Model Training: Train independent base models (e.g., decision trees, random forests, or gradient boosting machines) on each perturbed dataset, capturing diverse data patterns due to the introduced variability.

Step 3: Analytical Combination: Compute optimal combination weights for the base model predictions analytically, typically by solving optimization problems to minimize loss functions or deriving closed-form expressions.

Step 4: Final Prediction: Calculate the final prediction for a given input sample as the weighted sum of the base model predictions, using the optimized weights.

Step 5: Evaluation and Tuning: Assess the ensemble model's performance using appropriate metrics and fine-tune hyperparameters to optimize predictive accuracy.

Bagging

Bagging, proposed by (Breiman, 1996), is an ensemble learning technique that enhances the stability and accuracy of predictive models while mitigating overfitting. By training models on diverse subsets of the training data, Bagging reduces variance and improves robustness, particularly for weak learners such as decision trees. The Bagging algorithm operates as follows:

Step 1: Bootstrap Sampling: Generate multiple subsets of the original training dataset through bootstrap sampling, where each subset is of equal size but may include repeated samples and exclude others.

Step 2: Model Training: Train a predictive model on each bootstrap subset, enabling each model to learn distinct data characteristics due to the variability in the subsets.

Step 3: Prediction Aggregation: Combine predictions from all trained models, typically using majority voting for classification tasks to determine the final prediction.

Cascade Generalization

Cascade Generalization (CG), introduced by (Gama and Brazdil, 2000), is an ensemble learning method that enhances classification performance through sequential stacking. CG augments the original dataset with new attributes derived from the probability outputs of base models, reducing bias in attribute evaluation and improving predictive accuracy. Widely applied in natural disaster assessment (Chen et al., 2019; Pham et al., 2019) , CG operates as follows:

Step 1: Sequential Attribute Augmentation: Train a base classifier on the original dataset and append its probability outputs as new attributes to the dataset.

Step 2: Iterative Classifier Training: Sequentially train additional classifiers on the augmented dataset, with each classifier leveraging the enhanced feature set.

Step 3: Final Prediction: Combine predictions from the sequential classifiers to produce the final output, typically via weighted voting or averaging.

Dagging

Dagging, proposed by Ting and Witte (Ting and Witten, 1997), is an ensemble learning technique for classification tasks that improves model generalization by training sub-models on disjoint subsets of the dataset. This approach mitigates overfitting and enhances predictive performance. The Dagging algorithm proceeds as follows:

Step 1: Dataset Decomposition: Divide the training dataset into disjoint subsets using methods such as linear decomposition, scalar decomposition, or neural network-based decomposition.

Step 2: Sub-Model Training: Train independent predictive models on each subset, with each model addressing a specific segment of the data.

Step 3: Prediction Combination: Aggregate predictions from all sub-models, typically using majority voting, to generate the final classification output.

3.3. VALIDATION PARAMETERS

The models used for landslide susceptibility prediction are validated using evaluation metrics for classification problems (Bien et al., 2023; Le Minh et al., 2023; Pham et al., 2022; Wardhani et al., 2019), including AUC, Positive Predictive Value (PPV), Negative Predictive Value (NPV), Sensitivity (SST), Specificity (SPF), Accuracy (ACC), Kappa index, and Root Mean Square Error (RMSE). In there, AUC is a critical metric frequently utilized to evaluate the performance of classifiers (Chen and Chen, 2021). The AUC is determined by combining SST and SPF values at each predicted value threshold. The value of AUC ranges from 0 to 1, with a higher AUC indicating better model performance (Bien et al., 2023; Chen and Chen, 2021; Le Minh et al., 2023; Pham et al., 2022). The PPV, NPV, SST, SPF, and ACC metrics are expressed as percentages and are calculated based on four parameters derived from the confusion matrix. These parameters consist of True Positive (TP) and False Positive (FP), which respectively denote correctly and incorrectly predicted landslide samples; True Negative (TN) and False Negative (FN), representing correctly and incorrectly predicted non-landslide samples (Bien et al., 2023; Le Minh et al., 2023; Pham et al., 2022). Higher values of PPV, NPV, SST, SPF, and ACC, along with lower RMSE, indicate greater model accuracy (Dao et al., 2020). The Kappa index is used as a statistical measure of agreement between predicted and actual values (Baeza et al., 2016; Sterlacchini et al., 2011). The Kappa value ranges from 0 to 1, with a value closer to 1 indicating greater model accuracy (Prakash et al., 2024). A model is considered to have high confidence accuracy with Kappa > 0.59 (Prakash et al., 2024).

The formulas for calculating the metrics mentioned above are as follows (Le Minh et al., 2023):

$$PPV = TP/(TP + FP) \quad (1)$$

$$NPV = TN/(TN + FN \quad (2)$$

$$SST = TP/(TP + FN) \quad (3)$$

$$SPF = TN/(TN + FP) \quad (4)$$

$$ACC = (TP + TN)/(TN + FN + TP + FP) \quad (5)$$

$$RMSE = \sqrt{\frac{\sum(x_i - \hat{x}_i)^2}{N-P}} \quad (6)$$

where x_i and $\hat{x}_i$ are the actual and predicted landslide susceptibility values, and P is the number of estimated parameters, including the constant. N is the total number of landslide samples.

$$Kappa = \frac{P_0 - P_m}{1 - P_m} \quad (7)$$

where P_0 is the relative observed agreement among raters, and P_m is the assumed probability of random agreement.

3.4. EVALUATION ATTRIBUTE METHODS

Correlation Attribute Evaluation (CAE)

CAE employs the Pearson correlation coefficient to quantify the strength and direction of the linear relationship between two continuous variables (Nettleton, 2014). The value of the correlation coefficient lies between -1 and 1 (Nettleton, 2014). In this study, CAE assesses the correlation between each conditional factor and the binary landslide/non-landslide outcome in the training dataset (Lucchese et al., 2020). The correlation coefficient, ranging from - 1 to 1, is normalized to a 0 to 1 scale for consistency (Luat et al., 2020). A value close to 1 indicates a strong positive correlation, suggesting that the conditional factor significantly influences landslide occurrence, while a value near 0 denotes a weak linear relationship. The Pearson correlation coefficient is calculated as follows (Nettleton, 2014):

$$R = \left| \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2} \right| \quad (8)$$

where x_i and y_i are the values of the two variables. $\bar{x}$ and y are the means of the two variables.

Gain Ratio Attribute Evaluation (GRAE)

Gain Ratio Attribute Evaluation (GRAE) is a feature selection metric used in decision tree-based machine learning algorithms to identify the most informative attributes for classification tasks (Quinlan, 1986). GRAE quantifies the influence of conditional factors on landslide occurrence, with higher values indicating greater relevance. A GRAE value of 0 suggests no relationship between the factor and landslides. The GRAE calculation involves the following steps:

Step 1: Entropy Calculation: Entropy measures the impurity or randomness in the dataset based on the distribution of class labels (landslide/non-landslide). Higher entropy reflects greater disorder.

Step 2: Information Gain: This measures the reduction in entropy achieved by splitting the dataset based on a given attribute, indicating the attribute's contribution to data organization.

Step 3: Split Information: Split information quantifies the intrinsic randomness of an attribute's value distribution. Attributes with many distinct values yield higher split information.

Step 4: Gain Ratio: The gain ratio normalizes information gain by dividing it by split information, balancing the attribute's informativeness against its intrinsic complexity.

OneR method

OneR method is a straightforward attribute selection technique used in machine learning and data mining to evaluate the relevance of attributes to a target variable (Holte, 1993). n this study, OneR ranks conditional factors based on their influence on landslide occurrence, with higher OneR values indicating greater predictive power (Le Minh et al., 2023). The OneR algorithm operates as follows:

Step 1: Target Variable Selection: Identify the target variable (landslide/non-landslide) to be predicted.

Step 2: Data Grouping by Attribute: For each attribute, group the dataset according to the attribute's values

Step 3: Identifying the Most Common Class: Within each group, determine the most frequent class (landslide or non-landslide) for the target variable.

Step 4: Rule Creation: Generate simple rules of the form "If attribute A has value X, predict class Y" based on the most common class in each group.

Step 5: Accuracy Assessment: Evaluate the predictive accuracy of the rules for each attribute.

Step 6: Attribute Selection: Select the attribute yielding the highest prediction accuracy as the OneR model for the dataset.

3.5. MATERIALS

Landslide inventories

Landslide inventory data, comprising both landslide and non-landslide locations, are critical for training and validating machine learning models (Gu et al., 2024; Tehrani et al., 2021; Yang et al., 2023). In this study, historical landslide data within the Ha Long – Van Don Highway study area were compiled through a two-step process: (1) digitization of landslide locations using high-resolution satellite imagery from Google Earth, and (2) verification via field surveys. A total of 77 landslide locations were identified and represented as polygons. To convert the data into a format understandable by machine learning models, landslide data was assigned a value of 1, and non-landslide data was assigned a value of 0. To facilitate machine learning analysis, these polygons were converted into a point-based format at a 10 m/pixel spatial resolution, resulting in 3,263 landslide points labeled with a value of 1. An equal number of non-landslide points (3,263), labeled with a value of 0, were sampled to maintain a 1:1 ratio. This balanced distribution is strategically chosen to mitigate the risk of model over-fitting toward the majority class and to ensure the optimal sensitivity of the DPCT-based ensemble models, a methodological standard supported by established landslide susceptibility literature. The non-landslide sampling followed a two-step procedure: (1) random selection of points from map layers with slopes less than 5° and curvature values between -0.05 and 0.05 (indicative of flat terrain), and (2) verification and normalization through field surveys. To ensure the robustness of the sampling process and minimize potential artifacts from a single random iteration, multiple random sampling sets were initially tested; however, as the variance in model accuracy remained negligible, a final representative set was utilized for the analysis. The dataset was split into a training set (70 %, 2,392 landslide points) and a validation set (30 %, 871 landslide points), with non-landslide points similarly divided.

Conditional factors

Landslides result from the complex interplay of conditional factors, including geological, topo-graphical, geomorphological, land cover, and meteorological characteristics (Le Minh et al., 2023). Among these, rainfall often acts as a triggering factor, while other factors predispose areas to landslide events (Crosta and Frattini, 2008; Polemio and Petrucci, 2000; Zhang et al., 2019). This study selected 14 conditional factors to develop landslide susceptibility maps for the Ha Long – Van Don Highway, based on expert knowledge, local conditions, and statistical evaluation methods (Correlation Attribute Evaluation, Gain Ratio Attribute Evaluation, and OneR). All factors were standardized to a 10 m spatial resolution to align with the Digital Elevation Model (DEM) and the high-resolution landslide inventory. For categorical and continuous variables with coarser original resolutions, such as "Geology" and "Rainfall," a bilinear interpolation resampling technique was employed to downscale the data to the 10 m grid. While this downscaling process may introduce minor spatial uncertainties, it is essential for capturing localized terrain variations critical for highway-corridor analysis. To minimize potential artifacts, the resampled layers were cross-validated with original field data and existing thematic maps to ensure the maintenance of geological and meteorological continuity. These factors are detailed in Table 2 and illustrated in Figure 4. Their roles in landslide susceptibility are described below:

1. Elevation (m) reflects terrain potential, with higher elevations typically associated with greater landslide susceptibility due to increased gravitational forces (Bien et al., 2023; Le Minh et al., 2023).
2. Weathering crust: This factor indicates the degree of rock degradation, influencing soil stability and landslide propensity (Mai, 1996; Phong et al., 2021; Thanh et al., 2020).
3. Geology: Geological characteristics, such as rock type and structure, indirectly affect soil and rock physical properties, impacting slope stability (Chacón et al., 2006; Ohlmacher, 2000; Sitányiová et al., 2015).
4. Geotechnical Engineering: This factor describes soil and rock mechanical properties, which are critical to slope stability (Sitányiová et al., 2015).

Table 2 The source of the conditional factors used in this study.

No	Variable	Scale	Source
1	Elevation (m)	10 m	DEM (generated from a 1:10,000 scale topographic map fro Department of Survey, Mapping, and Geographic Information Viet)
2	Weathering crust	10 m	Survey report on the natural project, code 105.08-2020.25
3	Geology	10 m	Geological map at scale 1:50,000 from General Department of Geolo Minerals of Viet Nam
4	Geotechnical Engineering	10 m	Survey report on the natural project, code 105.08-2020.25
5	Hydrogeology	10 m	Survey report on the natural project, code 105.08-2020.25
6	LULC	10 m	Survey report on the natural project, code 105.08-2020.25
7	Rainfall (mm/day)	10 m	Viet Nam Meteorological and Hydrological Administration
8	Fault density (km/km2)	10 m	Geological map
9	Stream density (km/km2)	10 m	Generate from DEM
10	Slope (degree)	10 m	Generate from DEM
11	Aspect	10 m	Generate from DEM
12	Curvature	10 m	Generate from DEM
13	TWI	10 m	Generate from DEM
14	SPI	10 m	Generate from DEM

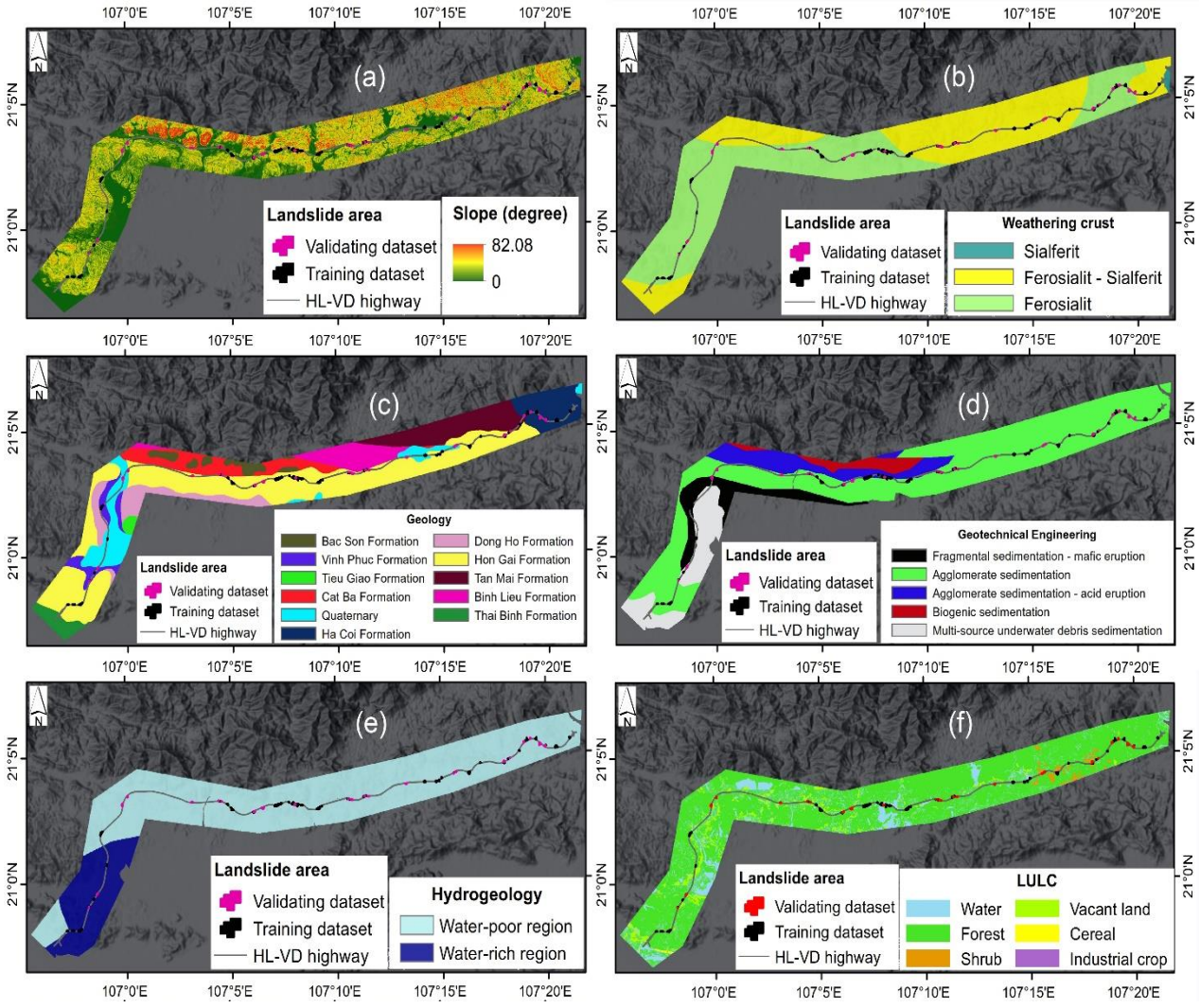

Fig. 4 Conditional factor maps for landslide susceptibility mapping in the Ha Long - Van Don Highway area.

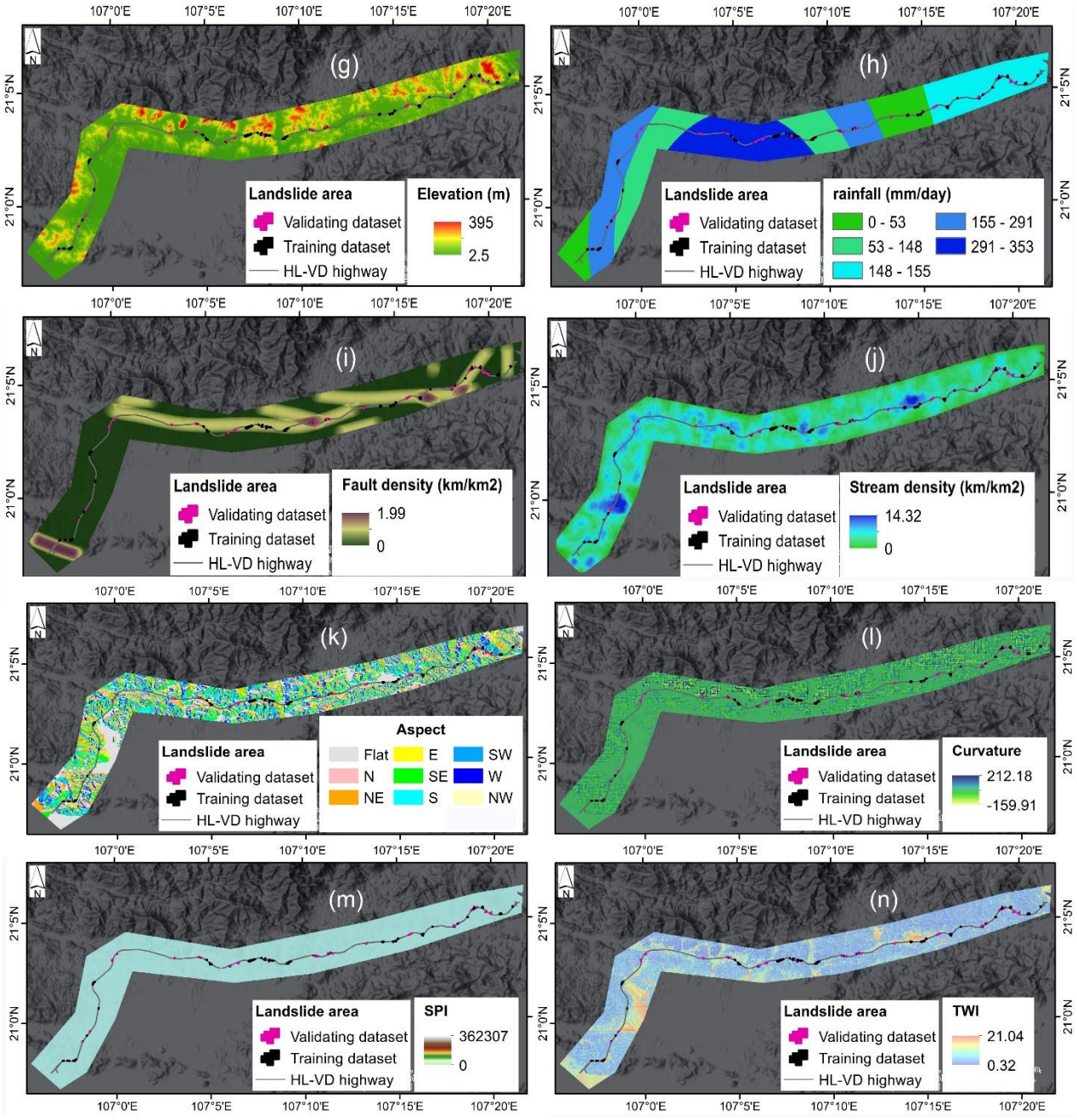

Fig. 4 Continued

5. Hydrogeological characteristics indirectly indicate the water retention capacity of soil and rocks, which is related to the conditions of water pressure in voids within the soil (Tacher et al., 2005).
6. Land use/Land cover (LULC) represents the vegetation cover characteristics on the land, typically, areas with dense forest cover have lower landslide probabilities (Rabby et al., 2022).
7. Rainfall amount (mm/day): As a primary trigger, rainfall infiltrates soil, increasing saturation and disrupting soil-rock cohesion. Higher daily rainfall averages, classified using the natural breaks method, correlate with increased landslide likelihood (Zhang et al., 2019).
8. Fault density (km/km2) is characteristic of the degree of rock destruction by tectonic; areas with higher fault densities experience greater rock destruction, facilitating landslides (Le Minh et al., 2023).
9. Stream density (km/km2) indirectly indicates the water retention capacity of soil and the drainage conditions on the terrain (Shirzadi et al., 2017). Typically, areas with higher flow densities are more favorable for landslides.
10. Slope (degree): is also an important factor for landslide occurrence (Çellek, 2020). Generally, slopes ranging from 25^0 to 40^0 are favorable for landslides (Çellek, 2020).

Table 3 The ranking conditional factors used in the selected methods.

Rank	Methods					
	CAE		Gain Ratio AE		OneR	
	Factor	*Values*	*Factor*	*Values*	*Factor*	*Values*
1	Slope (degree)	0.65	Slope (degree)	0.22	Elevation (m)	86.30
2	TWI	0.59	Curvature	0.15	Slope (degree)	85.56
3	Elevation (m)	0.50	Elevation (m)	0.14	Curvature	80.61
4	Aspect	0.45	TWI	0.13	TWI	78.50
5	Rainfall (mm/day)	0.28	Aspect	0.12	Aspect	77.01
6	Fault density (km/km2)	0.24	Geology	0.11	Geology	75.36
7	Stream density	0.22	Geotechnical Engineering	0.07	Fault density (km/km2)	66.31
8	Weathering crust	0.18	SPI	0.07	SPI	65.81
9	Geotechnical Engineering	0.14	Stream density	0.06	Rainfall (mm/day)	65.62
10	Hydrogeology	0.10	Fault density (km/km2)	0.04	Geotechnical Engineering	63.07
11	Geology	0.06	Hydrogeology	0.04	Weathering crust	60.58
12	LULC	0.02	Rainfall (mm/day)	0.04	Stream density	60.39
13	Curvature	0.01	LULC	0.03	LULC	59.29
14	SPI	0.01	Weathering crust	0.03	Hydrogeology	57.21

11. Aspect represents the characteristics of windward slopes, indirectly related to the soil's moisture absorption from humid air streams (Seda, 2021).
12. Curvature: characterizes the surface terrain, where flat terrain (values from -0.05 to 0.05) usually experiences fewer landslides, while concave (<-0.05) and convex (>0.05) terrains are more favorable for landslides (Phong et al., 2021).
13. The Topographic Wetness Index (TWI) indirectly indicates the moisture retention conditions of the terrain, related to the soil's water saturation. Higher TWI values indicate greater moisture retention capacity in the soil, and vice versa (Conoscenti et al., 2008).
14. The Stream Power Index (SPI): is characteristic of the energy of the terrain. Higher SPI values correspond to higher landslide probabilities (Yilmaz, 2009).

4. RESULTS

Conditional factor importance

The evaluation of conditional factor importance using Correlation Attribute Evaluation (CAE), Gain Ratio Attribute Evaluation (GRAE), and OneR methods reveals distinct rankings for each factor, as summarized in Table 3. Slope emerges as the most influential factor according to CAE and GRAE, while OneR ranks it second, with Elevation identified as the most influential factor. Elevation is ranked third by both CAE and GRAE. The Topographic Wetness Index (TWI) is the second most influential factor per CAE but ranks fourth in GRAE and OneR assessments. Aspect is ranked fourth by CAE and fifth by GRAE and OneR. Curvature, conversely, ranks second and third in GRAE and OneR, respectively, but is notably ranked thirteenth by CAE, highlighting methodological variability. Rainfall is the fifth most important factor according to CAE but ranks ninth and twelfth in OneR and GRAE, respectively. Geology is consistently ranked sixth by GRAE and OneR but eleventh by CAE. Fault Density is ranked sixth, seventh, and tenth by CAE, OneR, and GRAE, respectively. The remaining factors—Weathering Crust, Geotechnical Engineering, Hydrogeology, Stream Power Index (SPI), and Land Use/Land Cover (LULC)—are consistently ranked as having lower influence across all methods. Collectively, Slope, Elevation, TWI, Aspect, Rainfall, Curvature, and Geology are identified as the primary drivers of landslide susceptibility in the study area.

Figure 5 illustrates the distribution of landslide and non-landslide points across the classes of each conditional factor, providing insights into their influence on landslide occurrence. Accordingly, for the Slope factor, non-landslide positions mainly concentrate in classes ranging from 0-50, particularly with class 0^0 having a predominant sample count of over 1700. Landslide positions are more evenly distributed across classes ranging from 0-45^0, with fewer occurrences in classes greater than 45^0. Regarding the Weathering crust, landslides and non-landslides are predominantly distributed across two classes: Ferosialit and Ferosialit–Sialferit. For the Geology factor, landslides are predominantly concentrated in the Hon Gai Formation (≈ 2200 samples), Ha Coi Formation (≈ 630 samples), and Quaternary (≈ 250 samples). Conversely, non-landslide positions are more evenly distributed across predominant classes, particularly in the Hon Gai formation (≈ 1050 samples) and Quaternary (≈ 650 samples). Continuing with the Geotechnical Engineering factor, both landslides and non-landslides are concentrated on class G2, with sample counts of approximately 2450 and 2000, respectively. Similarly, for the Hydrogeology factor, landslide and non-landslide positions are predominantly concentrated on the Water-poor region class, with sample counts of 3000 and 2500, respectively. For LULC, landslides

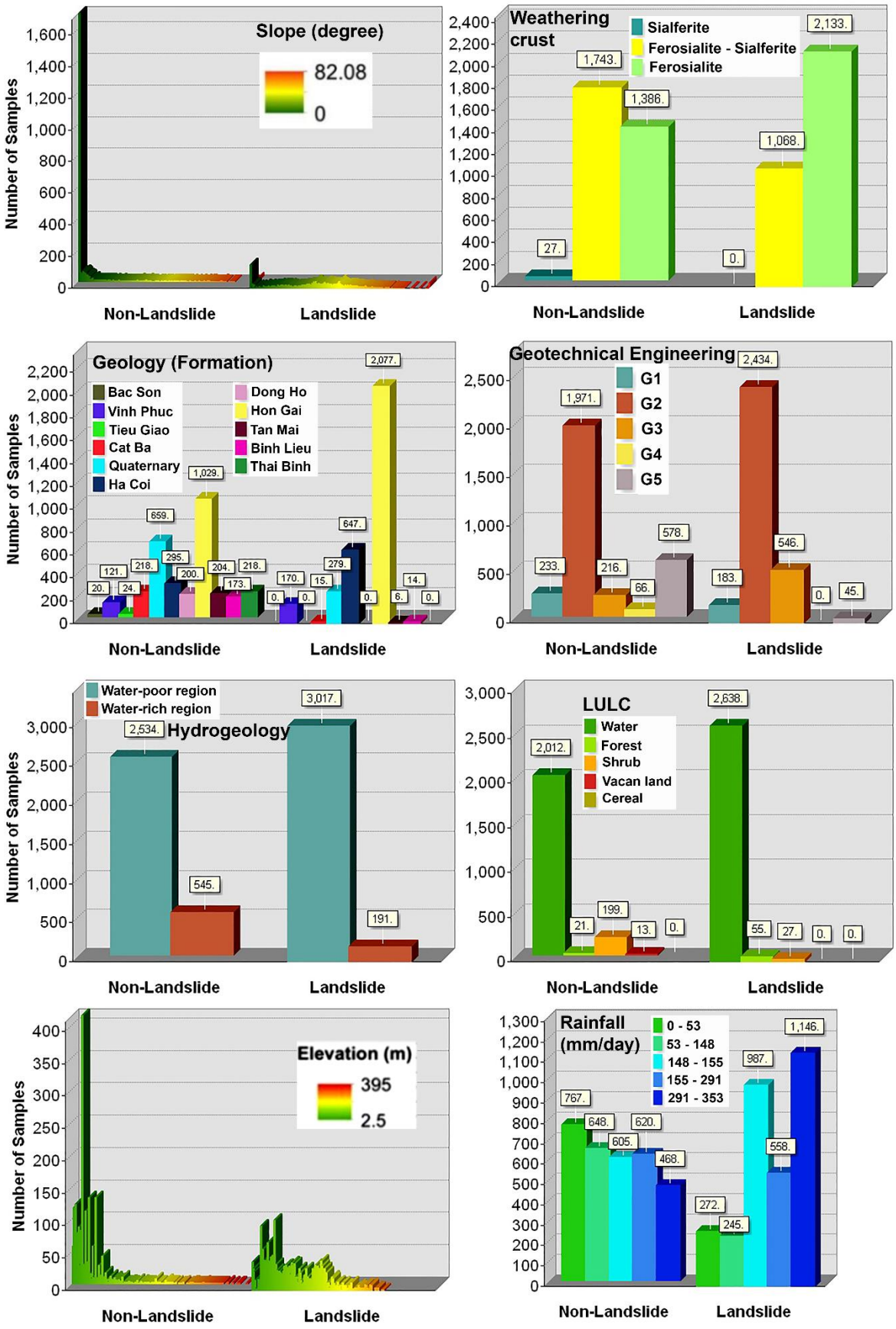

Fig. 5 Zonal histograms between landslide and non-landslide inventories in the study.

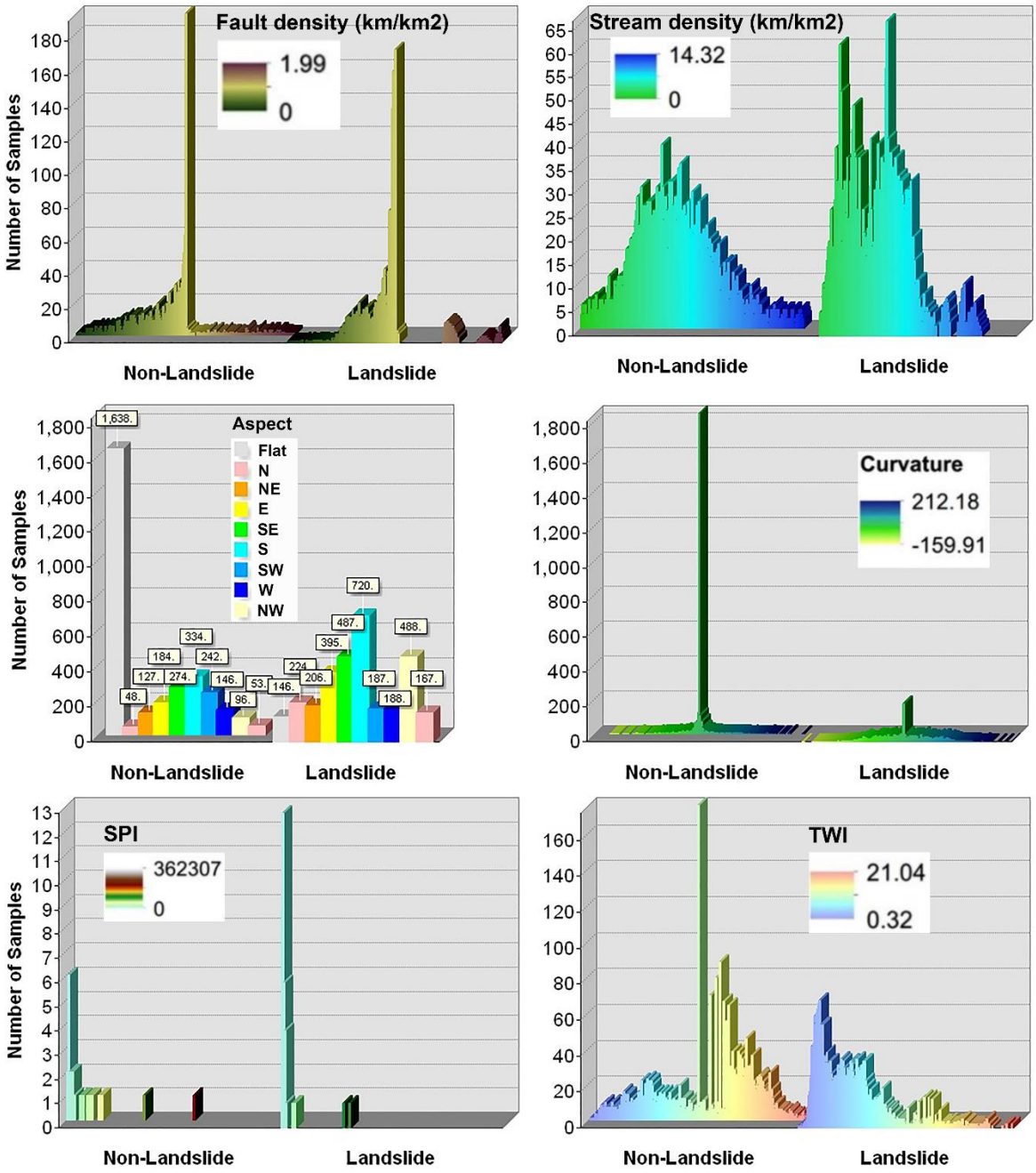

Fig. 5 Continued.

and non-landslides are predominantly distributed in the Forest class, with sample counts of approximately 2600 and 2000, respectively. Concerning Elevation, landslides are primarily distributed at elevations ranging from 2.5 to 250 m, while non-landslides are distributed at lower elevations (< 50 m). Subsequently, Rainfall, landslides, and non-landslides are distributed across all classes. Landslide positions are concentrated in classes with high rainfall (291 – 353 mm/day) and moderate rainfall (148 – 155 mm/day), with sample counts of approximately 1150 and 1000, respectively. Non-landslide positions are concentrated in classes with low rainfall (0 – 148 mm/day), with approximately 1400 samples. Landslides and non-landslides are distributed across classes ranging from 0 – 1 km/km^2 for the Fault density factor, with the highest concentration in the class with a value of 1 km/km^2. Regarding Stream density, landslide positions concentrate from 0 – 11 km/km^2, while non-landslide positions predominantly concentrate from 5- 12 km/km^2. Next, for the Aspect factor, landslides are fairly evenly distributed across directions, with a major concentration in the South (S) class with 700 samples. Non-landslide positions predominantly concentrate in the Flat class with approximately 1700 samples. Considering the Curvature factor, landslide positions are scattered across the value range, whereas non-landslide positions predominantly concentrate within the value range from -0.05 to 0.05. For the SPI factor, landslides and non-landslides are sporadically

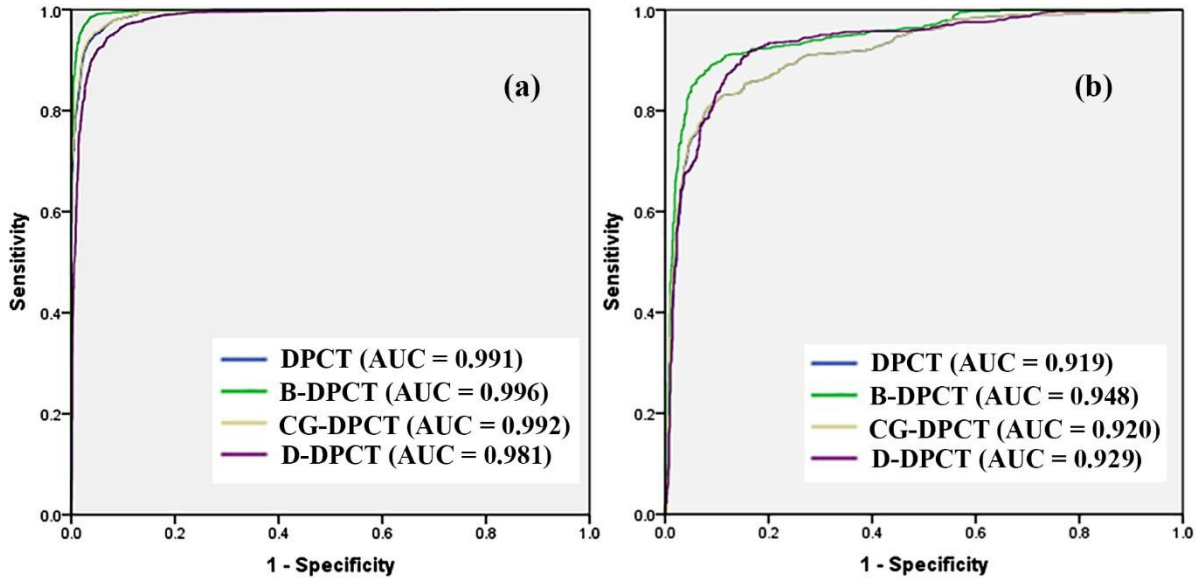

Fig. 6 AUC performance of the models: a) Training dataset, b) Validation dataset.

Table 4 Performance of the selected models.

No	Parameters	Models							
		Training dataset				Validating dataset			
		DPCT	B-DPCT	CG-DPCT	D-DPCT	DPCT	B-DPCT	CG-DPCT	D-DPCT
1	TP	2220	2221	2219	2198	948	949	948	946
2	TN	1820	2113	1872	2155	495	591	521	569
3	FP	16	15	17	38	23	22	23	25
4	FN	572	279	520	237	376	280	350	302
5	PPV (%)	99.28	99.33	99.24	98.30	97.63	97.73	97.63	97.43
6	NPV (%)	76.09	88.34	78.26	90.09	56.83	67.85	59.82	65.33
7	SST (%)	79.51	88.84	81.01	90.27	71.60	77.22	73.04	75.80
8	SPF (%)	99.13	99.30	99.10	98.27	95.56	96.41	95.77	95.79
9	ACC (%)	87.29	93.65	88.40	94.06	78.34	83.60	79.75	82.25
10	Kappa	0.75	0.87	0.77	0.88	0.56	0.67	0.59	0.64
11	RMSE	0.30	0.30	0.30	0.29	0.40	0.37	0.39	0.37

and unevenly distributed across classes. Finally, for the TWI factor, landslide positions predominantly concentrate in the value range from 0.32 – 8.0, while non-landslide positions predominantly concentrate in the value range from 10 - 21.

Evaluation of models

The reliability and accuracy of the models based on key parameters, including AUC, PPV (%), NPV (%), SST (%), SPF (%), ACC (%), Kappa, and RMSE, are evaluated. The numerical evaluation results of the models are detailed in Figure 6 and Table 4. Accordingly, the B-DPCT model yields the best results on the validation set with AUC = 0.948, PPV = 97.73 %, NPV = 67.85 %, SST = 73.4 %, SPF = 96.42 %, ACC = 83.6 %, Kappa = 0.67, and RMSE = 0.39 (Fig. 6b, Table 4).

Landslide susceptibility maps

Using all the datasets for the entire study area, we have established four landslide susceptibility maps created by four models, namely DPCT, B-DPCT, CG-DPCT, and D-DPCT (Fig. 7). Each map is divided into five susceptibility classes: very low, low, moderate, high, and very high (Fig. 7) using the natural break method. According to the susceptibility classes, we found that the single model DPCT has a high area ratio in the very low (41.3 %) and low (41.3 %) classes, while the moderate, high, and very high classes only account for 17.6 % (Figs. 7a, 8a). The CG-DPCT model is similar to DPCT, with the majority of the area in the very low and low classes (56.8 %), while the remaining three classes account for 43.2 % (Figs. 7c, 8a). Likewise, the B-DPCT model dominates the area in the very low and low classes

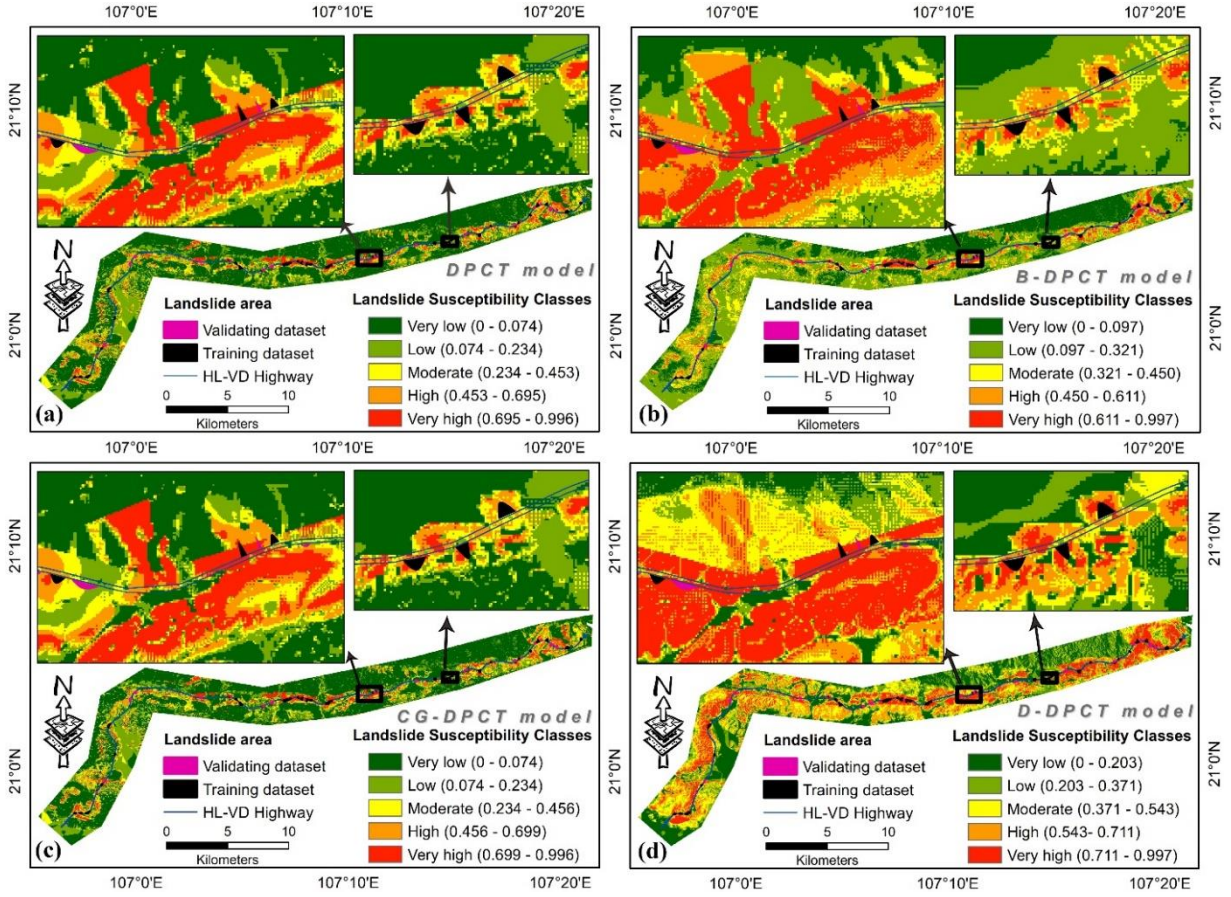

Fig. 7 Landslide susceptibility maps for the Ha Long – Van Dong highway: a) DPCT, b) B-DPCT, c) CG-DPCT, d) D-DPCT.

(54.1 %), with the other three classes accounting for 45.9 % (Figs. 7b, 8a). In the D-DPCT model, the landslide susceptibility classes are evenly distributed: very low (18.8 %), low (14.7 %), moderate (15.8 %), high (14.1 %), and very high (10.8 %) (Figs. 7d, 8a).

5. DISCUSSION

Recent studies have emphasized the evaluation of machine learning model performance using validation datasets to assess their effectiveness in LSM (Bien et al., 2023; Bui et al., 2020; Dao et al., 2020; Ghasemian et al., 2020; Le Minh et al., 2023; Nhu et al., 2022; Pham et al., 2022; Phong et al., 2021; Shahzad et al., 2022; Thanh et al., 2020). The evaluation results based on metrics of the base DPCT model are as follows: AUC = 0.919, PPV = 97.63 %, NPV = 56.83 %, SST = 71.60 %, SPF = 95.56 %, ACC = 78.34 %, Kappa = 0.56, and RMSE = 0.4. Consequently, the Bagging ensemble technique enhances the performance of the DPCT model with the following evaluation metrics: AUC = 0.948 (increase of 0.029), PPV = 97.73 % (increase of 0.1 %), NPV = 67.85 % (increase of 11.02 %), SST = 77.22 % (increase of 5.62 %), SPF = 96.41 % (increase of 0.85 %), ACC = 83.60 % (increase of 5.26 %), Kappa = 0.67 (increase of 0.11), RMSE = 0.37 (decrease of 0.3). Similarly, the Cascade Generalization technique also improves the performance of the base DPCT model with the following evaluation metrics: AUC = 0.920 (increase of 0.001), PPV = 97.63 % (unchanged), NPV = 59.82 % (increase of 2.99 %), SST = 73.04 % (increase of 1.44 %), SPF = 95.77 % (increase of 0.21 %), ACC = 79.75 % (increase of 1.41 %), Kappa = 0.59 (increase of 0.03), RMSE = 0.39 (decrease of 0.1). Lastly, the Dagging technique also improves the performance of the base DPCT model with the following evaluation metrics: AUC = 0.932 (increase of 0.013), PPV = 97.43 % (decrease of 0.2 %), NPV = 65.33 % (increase of 8.5 %), SST = 75.80 % (increase of 4. 2 %), SPF = 95.79 % (increase of 0.23 %), ACC = 82.25 % (increase of 3.91 %), Kappa = 0.64 (increase of 0.08), RMSE = 0.37 (decrease of 0.3). The analysis results above demonstrate that the Bagging, Cascade Generalization, and Dagging techniques all have the potential to enhance the performance of the DPCT model (Fig. 6, Table 4). In other landslide evaluation studies, these techniques also demonstrate the ability to improve the performance of a single model (Ali et al., 2024; Gu et al., 2024; Zhao et al., 2024). This indicates that ensemble learning techniques can easily enhance the performance of the original single model

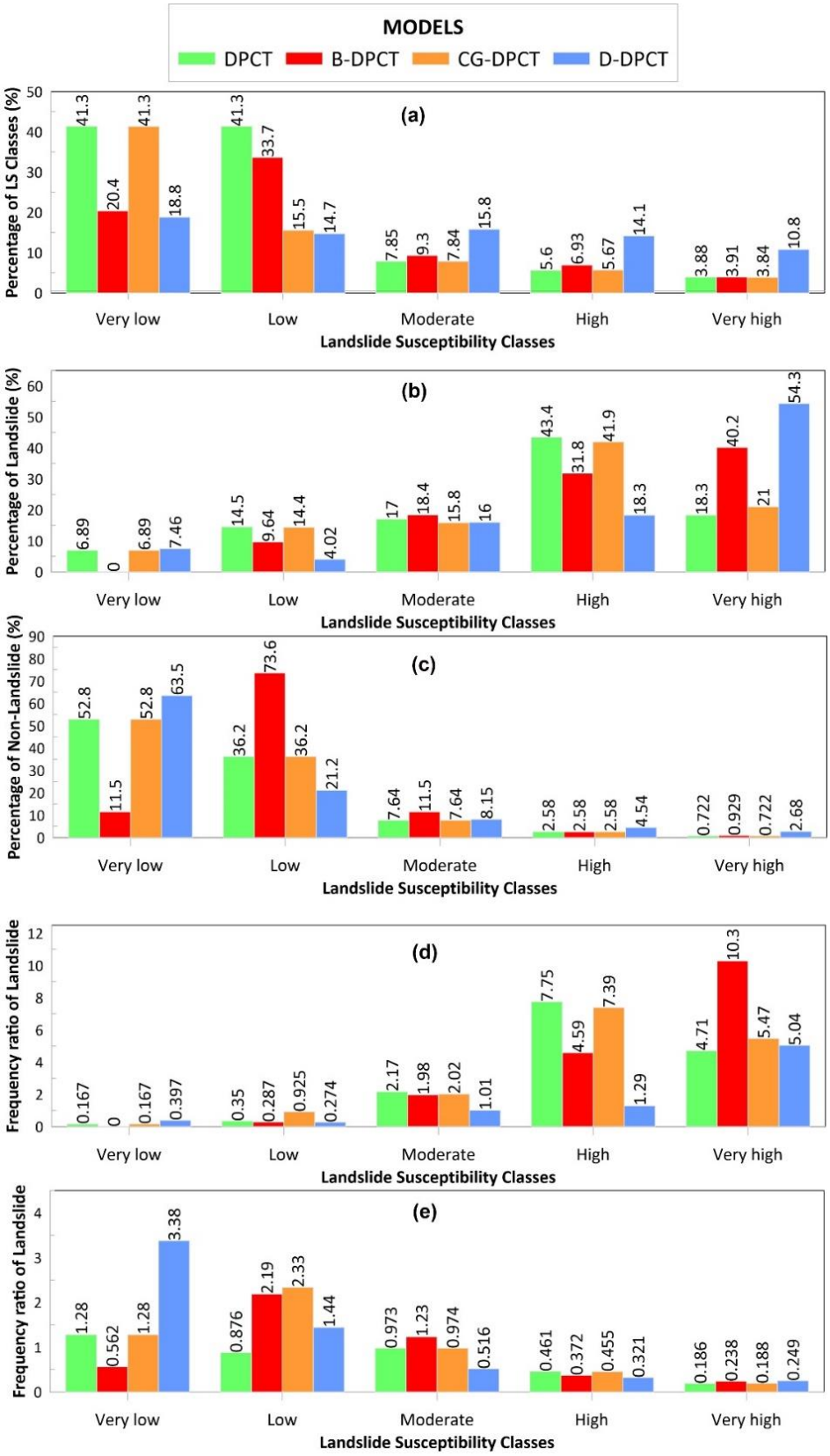

Fig. 8 Analysis results of landslide susceptibility maps:
a) Percentage of area of landslide susceptibility classes on each landslide susceptibility class, b) Percentage of validation landslide dataset on each landslide susceptibility class, c) Percentage of validation non-landslide dataset on each landslide susceptibility class, d) Frequency ratio of landslides on each landslide susceptibility class, and e) Frequency ratio of non-landslides on each landslide susceptibility class

(Liu et al., 2024; Singh et al., 2024). In this study, the Bagging ensemble based on the DPCT model exhibited the best performance compared to the other models.

Beyond performance metrics, the analysis of landslide susceptibility maps provides critical insights into model suitability for the Ha Long – Van Don Highway study area (Dao et al., 2020; Le Minh et al., 2023; Phong et al., 2021). Figure 8 illustrates the distribution of landslide and non-landslide points across susceptibility classes (very low to very high). Across all models, the percentage of landslides and their frequency ratio (FR) increase from very low to very high susceptibility classes, with landslides predominantly concentrated in high and very high classes (Figs. 8b, 8d). Conversely, the percentage of non-landslides and their FR decrease from very low to very high classes, with non-landslide points primarily located in very low and low classes (Figs. 8c, 8e). According to the FR evaluation method, higher FR values in high and very high landslide susceptibility classes and lower FR values in very low and low non-landslide classes indicate greater map reliability (Bien et al., 2023; Dao et al., 2020; Le Minh et al., 2023; Phong et al., 2021). Conversely, the lower the FR value of non-landslide classes lies below the low and very low landslide susceptibility classes, the more reliable the landslide susceptibility map becomes. The B-DPCT model yields the highest FR for landslides in the high (FR = 4.59) and very high (FR = 10.03) classes (Fig. 8d), while the D-DPCT model performs best for non-landslides in the very low (FR = 3.38) and low (FR = 1.44) classes (Fig. 8e), reinforcing the superior reliability of B-DPCT for LSM in this context.

The analysis of the 14 conditional factors using CAE, GRAE, and OneR methods reveals varying importance rankings (Table 3), indicating that each factor contributes uniquely to landslide occurrence (Keshri et al., 2023; Yu, et al., 2024). Notably, Slope, Elevation, and TWI consistently emerged as the most dominant drivers of landslide occurrence. The high ranking of the "Slope" factor (ranging from 25° to 40°) is directly correlated with the extensive cut-slope construction practices along the Ha Long – Van Don Highway corridor. As highlighted in the geotechnical assessments by Nguyen et al. (2020), the excavation of these steep engineered slopes has disturbed the natural hillslope equilibrium, making them highly susceptible to failure during intense monsoonal rainfall. Similarly, Elevation and TWI reflect the geomorphological reality of the area; higher elevations often coincide with the mountainous segments of the highway where deeper weathering crusts are prevalent, while higher TWI values indicate zones of significant moisture accumulation, which facilitates the reduction of soil shear strength and triggers instability.

Unlike other studies that include factors such as distance to roads or traffic density (Dao et al., 2020; Le Minh et al., 2023; Phong et al., 2021), this study excluded road-related factors to avoid introducing noise, given the focus on the Ha Long – Van Don Highway. Since the entire study area is essentially a 500 m-1000 m buffer zone of the highway corridor, "distance to road" would be a redundant parameter. Instead, focusing on geomorphological factors like Slope and TWI provides a more granular understanding of why specific sections of the highway are more hazardous than others.

The low NPV results (Table 4) observed across all models further underscore the sensitivity of machine learning to non-landslide sampling. While a 1:1 ratio was maintained to ensure a balanced training environment, future research should explore advanced sampling strategies to better distinguish between truly stable ground and zones that appear stable but possess latent geomorphological (Gu et al., 2024; Yang et al., 2023). In subsequent studies, more attention should be paid to evaluating the selection of non-landslide locations (Gu et al., 2024; Huang et al., 2024; Yang et al., 2023). Additionally, we believe that the performance of the models in future studies can be improved by the following enhancements: 1) adjusting the structural parameters of the model, selecting the optimal model structure, 2) enhancing the detail and resolution of the conditioning factor maps, 3) ensuring consistency in the research level and scale of the data, 4) considering weighting the predicted labels according to the scale of each landslide location.

The Ha Long - Van Don Highway has been in operation for just over six years (Nguyen, Tien and Do, 2020). Despite its relatively recent inauguration, landslides have occurred annually along the route, particularly during periods of intense rainfall, and the potential for future occurrences remains substantial. Consequently, the application of advanced techniques and the enhancement of landslide susceptibility mapping (LSM) resolution are both urgent and necessary. These improvements are expected to significantly increase the accuracy of landslide hazard prediction not only within the study area but also in other areas of interest. To date, limited research has been conducted to assess landslide hazards in this region. Among the few notable contributions, Nguyen et al. (2020) and Van Tien et al. (2021) primarily investigated the current status and rainfall-induced triggers of landslides (Nguyen et al., 2020; Van Tien et al., 2021). The present study builds in part upon the findings of Luat et al. (2024) but differs substantially in both methodological approach and data detail (Luat et al., 2024). Specifically, the landslide inventory has been updated with higher spatial accuracy, and a more comprehensive set of 14 influencing factors has been selected as model inputs. In the work of Luat et al. (2024), LSM for the Ha Long – Van Don Highway was produced through a comparative analysis of the Analytical Hierarchy Process (AHP) and Frequency Ratio (FR) methods (Luat et al., 2024). While these approaches are widely used, they share a major drawback: a heavy reliance on expert judgment, which can introduce subjectivity. By contrast, machine learning methods autonomously determine factor weights and decision rules. Leveraging their capacity

for generalization and processing large-scale datasets, machine learning models provide a more objective and potentially more accurate alternative to traditional approaches.

There are still many issues related to improving the performance of models, such as selecting non-landslide locations, choosing conditional factors, the level of detail of the data, and the structure of the selected models. However, this is the first study to successfully apply combined models to enhance the performance of DPCT models in establishing landslide susceptibility maps in highway areas. We recommend applying the Bagging ensemble based on the DPCT model in establishing landslide susceptibility maps for similar conditions. The landslide susceptibility map based on the B-DPCT model in the research area is recommended for use in management, construction planning, and disaster prevention and mitigation efforts. The classes of the landslide susceptibility map provide a solid scientific basis for managers to carry out the aforementioned tasks.

6. CONCLUSIONS

This study successfully implemented a novel DPCT-based ensemble framework to map landslide susceptibility along the Ha Long – Van Don Highway. The integration of Bagging with the DPCT base-learner (B-DPCT) yielded the most robust results (AUC = 0.948, ACC = 83.60 %, 83.60 %, Kappa = 0.67, and RMSE = 0.37), significantly outperforming individual models and traditional heuristic approaches like AHP. The core findings indicate that approximately 10 % of the highway corridor falls within high to very high susceptibility zones. These areas are characterized by steep engineered cut-slopes and high topographic moisture retention, confirming that human-induced terrain modification is a primary driver of risk in this region. Methodologically, the use of high-resolution (10 m) data proved essential in identifying localized hazards that coarser models would likely overlook.

The analysis of 14 conditional factors using Correlation Attribute Evaluation (CAE), Gain Ratio Attribute Evaluation (GRAE), and OneR methods revealed varying degrees of influence, with Slope, Elevation, Aspect, Topographic Wetness Index (TWI), Rainfall, Curvature, and Geology identified as the most significant drivers of landslide occurrence. These findings underscore the importance of tailored factor selection to reflect local geoenvironmental conditions.

The B-DPCT model identified approximately 10% of the study area as having high to very high landslide susceptibility, predominantly along the highway corridor. The resulting susceptibility maps provide a robust scientific foundation for governmental agencies to inform land-use planning, infrastructure development, and disaster mitigation strategies, thereby reducing landslide-related risks.

Despite these achievements, the study is limited by the inherent uncertainties of downscaling rainfall and geological data. Future efforts should focus on integrating real-time geotechnical monitoring data and optimizing model architectures to further refine prediction accuracy. Ultimately, the susceptibility maps produced herein serve as a vital decision-support tool for provincial authorities in managing infrastructure safety and implementing timely disaster mitigation strategies along this strategic transportation route.

FUNDING

This research was funded by the Vietnam National Foundation for Science and Technology Development (NAFOSTED) under grant number 105.08-2020.25.

CONFLICT OF INTEREST

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest

REFERENCES

Ado, M., Amitab, K., Maji, A.K., Jasińska, E., Gono, R., Leonowicz, Z. and Jasiński, M.: 2022, Landslide susceptibility mapping using machine learning: A literature survey. Remote Sens., 14, 3029. DOI: 10.3390/rs14133029

AghaKouchak, A., Chiang, F., Huning, L.S., Love, C.A., Mallakpour, I., Mazdiyasni, O., Moftakhari, H., Papalexiou, S.M., Ragno, E. and Sadegh, M.: 2020, Climate extremes and compound hazards in a warming world. Annu. Rev. Earth Planet. Sci., 48, 519–548. DOI: 10.1146/annurev-earth-071719-055228

Ajin, R.S., Saha, S., Saha, A., Biju, A., Costache, R. and Kuriakose, S.L.: 2022, Enhancing the accuracy of the REPTree by integrating the hybrid ensemble meta-classifiers for modelling the landslide susceptibility of Idukki District, South-western India. J. Indian Soc. Remote Sens., 50, 2245–2265. DOI: 10.1007/s12524-022-01599-4

Ali, N., Chen, J., Fu, X., Ali, R., Hussain, M.A., Daud, H., Hussain, J., and Altalbe, A.: 2024, Integrating machine learning ensembles for landslide susceptibility mapping in Northern Pakistan. Remote Sens., 16, 988. DOI: 10.3390/rs16060988

Asadi, M., Goli Mokhtari, L., Shirzadi, A., Shahabi, H. and Bahrami, S.: 2022, A comparison study on the quantitative statistical methods for spatial prediction of shallow landslides (Case study: Yozidar-Degaga Route in Kurdistan Province, Iran). Environ. Earth Sci., 81, 51. DOI: 10.1007/s12665-021-10152-4

Azarafza, M., Azarafza, M., Akgün, H., Atkinson, P.M. and Derakhshani, R.: 2021, Deep learning-based landslide susceptibility mapping. Sci. Rep., 11, 24112.

Azevedo, B.F., Rocha, A.M.A.C. and Pereira, A.I.: 2024, Hybrid approaches to optimization and machine learning methods: a systematic literature review. Mach. Learn, 113, 4055–4097. DOI: 10.1007/s10994-023-06467-x

Baeza, C., Lantada, N. and Amorim, S.: 2016, Statistical and spatial analysis of landslide susceptibility maps with different classification systems. Environ. Earth Sci., 75, 1318. DOI: 10.1007/s12665-016-6124-1

Bien, T.X., Iqbal, M., Jamal, A., Nguyen, D.D., Van Phong, T., Costache, R., Ho, L.S., Van Le, H., Nguyen, H.B.T., Prakash, I. and Pham, B.T.: 2023, Integration of rotation forest and multiboost ensemble methods with forest by penalizing attributes for spatial prediction of landslide susceptible areas. Stoch. Environ. Res. Risk Assess., 37, 4641–4660. DOI: 10.1007/s00477-023-02521-1

Breiman, L.: 1996, Bagging predictors. Mach. Learn., 24, 123–140. DOI: 10.1007/BF00058655

Bui, D.T., Tsangaratos, P., Nguyen, V.-T., Liem, N.V. and Trinh, P.T.: 2020, Comparing the prediction performance of a Deep Learning Neural Network model with conventional machine learning models in landslide susceptibility assessment. Catena, 188, 104426. DOI: 10.1016/j.catena.2019.104426

Bui, Q.D., Ha, H., Khuc, D.T., Nguyen, D.Q., von Meding, J., Nguyen, L.P. and Luu, C.: 2023, Landslide susceptibility prediction mapping with advanced ensemble models: Son La province, Vietnam. Nat. Hazards, 116, 2283–2309. DOI: 10 .21203/rs.3.rs-1650275/v1

Cao, W.G., Pan, D., Xu, Z.J., Zhang, W.P., Ren, Y. and Nan, T.: 2025. Landslide disaster vulnerability mapping study in Henan Province: Comparison of different machine learning models. Bull. Geol. Sci. Technol., 44, 1,101–111. DOI: 10.19509/j.cnki.dzkq.tb20230338

Çellek, S.: 2020, Effect of the slope angle and its classification on landslide. Nat. Hazards Earth Syst. Sci., 1–23. DOI: 10.5194/nhess-2020-87

Chacón, J., Irigaray, C., Fernández, T. and El Hamdouni, R.: 2006, Engineering geology maps: landslides and geographical information systems. Bull. Eng. Geol. Environ., 65, 341–411. DOI: 10.1007/s10064-006-0064-z

Chen, W., Yan, X., Zhao, Z., Hong, H., Bui, D. and Pradhan, B.: 2019, Spatial prediction of landslide susceptibility using data mining-based kernel logistic regression, naive Bayes and RBFNetwork models for the Long County area (China). Bull. Eng. Geol. Environ., 78, 247–266. DOI: 10.1007/s10064-018-1256-z

Chen, X. and Chen, W.: 2021, GIS-based landslide susceptibility assessment using optimized hybrid machine learning methods. Catena, 196, 104833. DOI: 10.1016/j.catena.2020.104833

Conoscenti, C., Di Maggio, C. and Rotigliano, E.: 2008, GIS analysis to assess landslide susceptibility in a fluvial basin of NW Sicily (Italy). Geomorphology, 94, 325–339. DOI: 10.1016/j.geomorph.2006.10.039

Costanzo, D. and Irigaray, C.: 2020, Comparing forward conditional analysis and forward logistic regression methods in a landslide susceptibility assessment: A case study in Sicily. Hydrology, 7, 37. DOI: 10.3390/hydrology7030037

Crosta, G.B. and Frattini, P.: 2008, Rainfall-induced landslides and debris flows. Hydrol. Process, 22, 473-477. DOI: 10.1002/hyp.6885

D'Amato, G. and Akdis, C.A.: 2020, Global warming, climate change, air pollution and allergies. Allergy, 75, 2158–2160. DOI: 10.1111/all.14527

Dao, D.V., Jaafari, A., Bayat, M., Mafi-Gholami, D., Qi, C., Moayedi, H., Phong, T.V., Ly, H.-B., Le, T.-T., Trinh, P.T., Luu, C., Quoc, N.K., Thanh, B.N. and Pham, B.T.: 2020, A spatially explicit deep learning neural network model for the prediction of landslide susceptibility. Catena, 188, 104451. DOI: 10.1016/j.catena.2019.104451

Demirel, B., Yildirim, E. and Can, E.: 2025, GIS-based landslide susceptibility mapping using AHP, FMEA, and Pareto systematic analysis in central Yalova, Türkiye. Eng. Sci. Technol., 64, 102013. DOI: 10.1016/j.jestch.2025.102013

Dey, S., Das, S. and Roy, S.K.: 2024, Landslide susceptibility assessment in eastern Himalayas, India: a comprehensive exploration of four novel hybrid ensemble data driven techniques integrating explainable artificial intelligence approach. Environ. Earth Sci., 83, 641. DOI: 10.1007/s12665-024-11945-z

Fang, Z., Wang, Y., Duan, G. and Peng, L.: 2021, Landslide susceptibility mapping using rotation forest ensemble technique with different decision trees in the Three Gorges Reservoir Area, China. Remote Sens.. 13, 238. DOI: 10.3390/rs13020238

Farinós-Dasí, J., Pinazo-Dallenbach, P., Peiró Sánchez-Manjavacas, E. and Rodríguez-Bernal, D.C.: 2024, Disaster risk management, climate change adaptation and the role of spatial and urban planning: evidence from European case studies. Nat. Hazards, 121, 23479–23512. DOI: 10.1007/s11069-024-06448-w

Gama, J. and Brazdil, P.: 2000, Cascade generalization. Mach. Learn., 41, 315–343. DOI: 10.1023/A:1007652114878

Geurts, P.: 2001, Dual perturb and combine algorithm. PMLR Proceedings of Machine Learning Research, R3, 106–111.

Geurts, P. and Wehenkel, L.: 2005, Closed-form dual perturb and combine for tree-based models. Proceedings of the 22nd international conference on Machine learning. Association for Computing Machinery, Bonn, Germany, 233–240. DOI: 10.1145/1102351.1102381

Ghasemian, B., Asl, D.T., Pham, B.T., Avand, M., Nguyen, H.D. and Janizadeh, S.: 2020, Shallow landslide susceptibility mapping: A comparison between classification and regression tree and reduced error pruning tree algorithms. Vietnam J. Earth Sci., 42, 208–227. DOI: 10.15625/0866-7187/42/3/14952

Ghiasi, V., Pauzi, N.I.M., Karimi, S. and Yousefi, M.: 2023, Landslide risk zoning using support vector machine algorithm. Geomech. Eng,. 34, 267–284. DOI: 10.12989/gae.2023.34.3.267

Gu, T., Duan, P., Wang, M., Li, J. and Zhang, Y.: 2024, Effects of non-landslide sampling strategies on machine learning models in landslide susceptibility mapping. Sci. Rep., 14, 7201. DOI: 10.1038/s41598-024-57964-5

He, Q., Wu, S., Zhao, X., Hui, Z., Wang, Z., Tsangaratos, P., Ilia, I., Chen, W., Chen, Y. and Hao, Y.: 2025, Evaluation of landslide susceptibility of mountain highway based on RF and SVM models. Sci. Rep., 15, 24991. DOI: 10.1038/s41598-025-08774-w

Holte, R.C.: 1993, Very simple classification rules perform well on most commonly used datasets. Mach. Learn., 11, 63–90. DOI: 10.1023/A:1022631118932

Hong, H.: 2023a, Assessing landslide susceptibility based on hybrid Best-first decision tree with ensemble learning model. Ecol. Indic., 147, 109968. DOI: 10.1016/j.ecolind.2023.109968

Hong, H.: 2023b, Assessing landslide susceptibility based on hybrid multilayer perceptron with ensemble learning. Bull. Eng. Geol. Environ., 82, 382. DOI: 10.1007/s10064-023-03409-8

Huang, F., Haowen, X., Jiang, S.-H., Yao, C., Fan, X., Catani, F., Chang, Z., Zhou, X., Huang, J. and Liu, K.: 2024, Modelling landslide susceptibility prediction: A review and construction of semi-supervised imbalanced theory. Earth-Sci. Rev., 250, 104700. DOI: 10.1016/j.earscirev.2024.104700

Ibrahim, M.B., Harahap, I.S.H., Balogun, A.-L.B. and Usman, A.: 2020, The use of geospatial data from GIS in the quantitative analysis of landslides. IOP Conference Series: Earth Environ. Sci., 540, 012048. DOI: 10.1088/1755-1315/540/1/012048

Kalantar, B., Ueda, N., Saeidi, V., Ahmadi, K., Halin, A.A. and Shabani, F.: 2020, Landslide susceptibility mapping: Machine and ensemble learning based on remote sensing big data. Remote Sens., 12, 1737. DOI: 10.3390/rs12111737

Keshri, D., Sarkar, K. and Chattoraj, S.L.: 2023, Landslide susceptibility mapping in parts of Aglar watershed, Lesser Himalaya based on frequency ratio method in GIS environment. J. Earth Syst. Sci., 133, 1. DOI: 10.1007/s12040-023-02204-z

Khosravi, K., Golkarian, A., Melesse, A.M. and Deo, R.C.: 2022, Suspended sediment load modeling using advanced hybrid rotation forest based elastic network approach. J. Hydrol., 610. DOI: 10.1016/j.jhydrol.2022.127963

Le Minh, N., Truyen, P.T., Van Phong, T., Jaafari, A., Amiri, M., Van Duong, N., Van Bien, N., Duc, D.M., Prakash, I. and Pham, B.T.: 2023, Ensemble models based on radial basis function network for landslide susceptibility mapping. Environ. Sci. Pollut. Res., 30, 99380–99398. DOI: 10.1007/s11356-023-29378-9

Li, X., Lu, Y., Fan, H., Xie, P., Yan, Y., Zhao, S., Lu, M. and Zhang, L.: 2025, Highway landslide susceptibility evaluation methods–a case study of Guangxi Zhuang autonomous region, China. Geomat. Nat. Hazards Risk, 161. DOI: 10.1080/19475705.2025.2563652

Liu, L.-L., Danish, A., Wang, X.-M. and Zhu W.-Q.: 2024, Ensemble stacking: a powerful tool for landslide susceptibility assessment – a case study in Anhua County, Hunan Province, China. Geocarto Int., 39, 2326005. DOI: 10.1080/10106049.2024.2326005

Liu, S., Wang, L., Zhang, W., He, Y., Pijush, S.: 2023, A comprehensive review of machine learning-based methods in landslide susceptibility mapping. Geol. J., 58, 2283–2301. DOI: 10.1002/gj.4666

Luat, N.-V., Do, T.-N., Nguyen, L.C. and Kien, N.T.: 2024, Assessing landslide susceptibility along the Halong-Vandon expressway in Quang Ninh province, Vietnam: A comprehensive approach integrating GIS and various methods. Geomech. Eng.,. 37, 135–147. DOI: 10.12989/gae.2024.37.2.135

Luat, N.-V., Nguyen, V.-Q., Lee, S., Woo, S. and Lee, K.: 2020, An evolutionary hybrid optimization of MARS model in predicting settlement of shallow foundations on sandy soils. Geomech. Eng., 21, 583–598. DOI: 10.12989/gae.2020.21.6.583

Lucchese, L.V., de Oliveira, G.G. and Pedrollo, O.C.: 2020, Attribute selection using correlations and principal components for artificial neural networks employment for landslide susceptibility assessment. Environ. Monit. Assess., 192, 129. DOI: 10.1007/s10661-019-7968-0

Mai, N.T.: 1996, Forecasting occurrence of landslide related to the tropical weathering crust by statistical analysis. Geoinformatics, 7, 91–95. DOI: 10 .6010/geoinformatics1990.7.1-2_91

Masson-Delmotte, V., Zhai, P., Pörtner, H.-O., Roberts, D., Skea, J. and Shukla, P.R.: 2022, Global warming of 1.5 °C. IPCC special report on impacts of global warming of 1.5 °C above pre-industrial levels in context of strengthening response to climate change, sustainable development, and efforts to eradicate poverty. Cambridge University Press. DOI: 10.1017/9781009157940

Nettleton, D.: 2014, Chapter 6 - Selection df variables and factor derivation. In: Nettleton, D. (Ed.), Commercial data mining. Morgan Kaufmann, Boston, 79–104. DOI: 10.1016/B978-0-12-416602-8.00006-6

Ngewie, D.T.L.: 2024, The impacts of road transport infrastructure and the socio-economic development in the Bamenda III Municipality, Mezam Division, North West Region Cameroon. Int. J. Bus. Dipl. Econ., 3. DOI: 10.51699/ijbde.v3i1.3322

Nguyen, L.C., Tien, P.V. and Do, T.-N.: 2020, Deep-seated rainfall-induced landslides on a new expressway: a case study in Vietnam. Landslides, 17, 395–407. DOI: 10.1007/s10346-019-01293-6

Nhu, V.-H., Bui, T.T., My, L.N., Vuong, H. and Duc, H.N.: 2022, A new approach based on integration of random subspace and C4.5 decision tree learning method for spatial prediction of shallow landslides. Vietnam J. Earth Sci., 44, 327–342. DOI: 10 .15625/2615-9783/16929

Ogunbode, C.A., Doran, R. and Böhm, G.: 2020, Exposure to the IPCC special report on 1.5 °C global warming is linked to perceived threat and increased concern about climate change. Clim. Change, 158, 361–375. DOI: 10.1007/s10584-019-02609-0

Ohlmacher, G.C.: 2000, The relationship between geology and landslide hazards of Atchison, Kansas, and vicinity. Curr. Res. Earth Sci., 1–16. DOI: 10.17161/cres.v0i244.11833

Pasang, S. and Kubíček, P.: 2020, Landslide susceptibility mapping using statistical methods along the Asian Highway, Bhutan. Geosciences, 10, 430. DOI: 10.3390/geosciences10110430

Pham, B., Prakash, I., Chen, W., Ly, H.-B., Ho, L., Omidvar, E., Tran, V. and Bui, D.: 2019, A novel intelligence approach of a sequential minimal optimization-based support vector machine for landslide susceptibility mapping. Sustainability, 11, 6323. DOI: 10.3390/su11226323

Pham, B.T., Vu, V.D., Costache, R., Phong, T.V., Ngo, T.Q., Tran, T.-H., Nguyen, H.D., Amiri, M., Tan, M.T., Trinh, P.T., Le, H.V. and Prakash, I.: 2022, Landslide susceptibility mapping using state-of-the-art machine learning ensembles. Geocarto Int., 37, 5175–5200. DOI: 10.1080/10106049.2021.1914746

Phi, F.G., Cho, B., Kim, J., Cho, H., Choo, Y.W., Kim, D. and Kim, I.: 2024, Machine learning application to seismic site classification prediction model using Horizontal-to-Vertical-Spectral Ratio (HVSR) of strong-ground motions. Geomech. Eng., 37, 539–554. DOI: 10.12989/gae.2024.37.6.539

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

Polemio, M. and Petrucci, O.: 2000, Rainfall as a landslide triggering factor an overview of recent international research. Thomas Telford Ltd., 1219-1226.

Prakash, I., Nguyen, D.D., Tuan, N.T. and Phong, T.V.: 2024, Landslide susceptibility zoning: Integrating multiple intelligent models with SHAP analysis. J. Sci. Transp. Technol., 4, 23–41. DOI: 10.58845/jstt.utt.2024.en.4.1.23-41

Quinlan, J.R.: 1986, Induction of decision trees. Mach. Learn. 1, 81–106. DOI: 10.1007/BF00116251

Rabby, Y.W., Li, Y., Abedin, J. and Sabrina, S.: 2022, Impact of land use/land cover change on landslide susceptibility in Rangamati Municipality of Rangamati District, Bangladesh. ISPRS Int. J. Geo-Inf., 11, 89. DOI: 10.3390/ijgi11020089

Ramos-Bernal, R.N., Vázquez-Jiménez, R., Cantú-Ramírez, C.A., Alarcón-Paredes, A., Alonso-Silverio, G.A., Bruzón, A.G., Arrogante-Funes, F., Martín-González, F., Novillo, C.J. and Arrogante-Funes, P.: 2021, Evaluation of conditioning factors of slope instability and continuous change maps in the generation of landslide inventory maps using Machine Learning (ML) algorithms. Remote Sens., 13, 4515. DOI: 10.3390/rs13224515

Samadi, H., Alanazi, A., Muhodir, S.H., Alsubai, S., Alqahtani, A. and Marzougui, M.: 2024, In-depth exploration of machine learning algorithms for predicting sidewall displacement in underground caverns. Geomech. Eng., 37, 307. DOI: 10.12989/gae.2024.37.4.307

Saha, A., Tripathi, L., Villuri, V.G.K. and Bhardwaj, A.: 2024, Exploring machine learning and statistical approach techniques for landslide susceptibility mapping in Siwalik Himalayan region using geospatial technology. Environ. Sci. Pollut. Res., 31, 10443–10459. DOI: 10.1007/s11356-023-31670-7

Saha, A., Villuri, V.G.K. and Bhardwaj, A.: 2025, Development and assessment of a novel hybrid machine learning-based landslide susceptibility mapping model in the Darjeeling Himalayas. Stoch. Environ. Res. Risk Assess., 39, 4145–4168. DOI: 10.1007/s00477-023-02528-8

Saha, A., Villuri, V.G.K., Bhardwaj, A., and Kumar, S.: 2023, A multi-criteria decision analysis (MCDA) approach for landslide susceptibility mapping of a part of darjeeling district in north-east Himalaya, India. Appl. Sci., 13, 5062. DOI: 10.3390/app13085062

Saha, A., Villuri, V.G.K. and Bhardwaj, A.: 2022, Development and assessment of GIS-based landslide susceptibility mapping models using ANN, Fuzzy-AHP, and MCDA in Darjeeling Himalayas, west Bengal, India. Land, 11, 10, 1711. DOI: 10.3390/land11101711

Sassa, K., Mikoš, M., Sassa, S., Bobrowsky, P.T., Takara, K. and Dang, K.: 2020, Understanding and reducing landslide disaster risk. Springer, Cham. DOI: 10.1007/978-3-030-60196-6

Seda, C.: 2021, The effect of aspect on landslide and its relationship with other parameters. In: Yuanzhi, Z., Qiuming, C. (Eds.), Landslides, chapter 1. IntechOpen. DOI: 10.5772/intechopen.99389

Shahzad, N., Ding, X. and Abbas, S.: 2022, A comparative assessment of machine learning models for landslide susceptibility mapping in the Rugged Terrain of Northern Pakistan. Appl. Sci., 12, 2280. DOI: 10.3390/app12052280

Shano, L., Raghuvanshi, T.K. and Meten, M.: 2020, Landslide susceptibility evaluation and hazard zonation techniques – a review. Geoenvironmental Disasters, 7, 18. DOI: 10.1186/s40677-020-00152-0

Shirzadi, A., Bui, D.T., Pham, B.T., Solaimani, K., Chapi, K., Kavian, A., Shahabi, H. and Revhaug, I.: 2017, Shallow landslide susceptibility assessment using a novel hybrid intelligence approach. Environ. Earth Sci., 76, 60. DOI: 10.1l07/s12665-016-6374-y

Singh, K., Bhardwaj, V., Sharma, A. and Thakur, S.: 2024, A comprehensive review on landslide susceptibility zonation techniques. Quaest. Geogr., 43. DOI: 10.14746/quageo-2024-0005

Singh, Y., Chib, S., Hamdani, I., Parihar, G.S., Johar, S. and Pandita, S.K.: 2025, Landslide susceptibility mapping along the Udhampur-Ramban section of National Highway-44, Jammu and Kashmir: A case study based on the frequency ratio method. Result. Earth Sci., 3, 100141. DOI: 10.1016/j.rines.2025.100141

Sitányiová, D., Vondráčková, T., Stopka, O., Myslivečková, M. and Muzik, J.: 2015, GIS based methodology for the geotechnical evaluation of landslide areas. Procedia Earth Planet. Sci., 15, 389–394. DOI: 10.1016/j.proeps.2015.08.011

Sterlacchini, S., Ballabio, C., Blahut, J., Masetti, M. and Sorichetta, A.: 2011, Spatial agreement of predicted patterns in landslide susceptibility maps. Geomorphology, 125, 51–61. DOI: 10.1016/j.geomorph.2010.09.004

Sun, D., Gu, Q., Wen, H., Xu, J., Zhang, Y., Shi, S., Xue, M. and Zhou, X.: 2023, Assessment of landslide susceptibility along mountain highways based on different machine learning algorithms and mapping units by hybrid factors screening and sample optimization. Gondwana Res., 123, 89–106. DOI: 10.1016/j.gr.2022.07.013

Tacher, L., Bonnard, C., Laloui, L. and Parriaux, A.: 2005, Modelling the behaviour of a large landslide with respect to hydrogeological and geomechanical parameter heterogeneity. Landslides, 2, 3–14. DOI: 10.1007/s10346-004-0038-9

Tang, H., Wang, C., An, S., Wang, Q. and Jiang, C.: 2023, A novel heterogeneous ensemble framework based on machine learning models for shallow landslide susceptibility mapping. Remote Sens., 15, 4159. DOI: 10.3390/rs15174159

Technology, V.I.f.B.S.a.: 2009, Vietnam Building Code Natural Physical and Climatic Data for Construction, Ministry of Constructions.

Tehrani, F.S., Santinelli, G. and Herrera, M.: 2021, Multi-regional landslide detection using combined unsupervised and supervised machine learning. Geomat. Nat. Hazards Risk, 12, 1015–1038. DOI: 10.1080/19475705.2021.1912196

Thanh, D.Q., Nguyen, D.H., Prakash, I., Jaafari, A., Nguyen, V.T., Phong, T.V. and Pham, B.T.: 2020, GIS based frequency ratio method for landslide susceptibility mapping at Da Lat City, Lam Dong province, Vietnam. Vietnam J. Earth Sci., 42, 55–66. DOI: 10.15625/0866-7187/42/1/14758

Thanh, T.-D.: 2011, Stratigraphic units of Vietnam (Second Edition - Revised and Updated). Vietnam National University Publisher, Hanoi, 556 pp.

Tin, D., Cheng, L., Le, D., Hata, R. and Ciottone, G.: 2024, Natural disasters: a comprehensive study using EMDAT database 1995–2022. Public Health, 226, 255–260. DOI: 10.1016/j.puhe.2023.11.017

Ting, K.M. and Witten, I.H.: 1997, Stacking bagged and dagged models. Fourteenth International Conference on Machine Learning, San Francisco, CA, 367–375. DOI: 10.5555/645526.657147

Tong, Z.l., Guan, Q.T., Arabameri, A., Loche, M. and Scaringi, G.: 2023, Application of novel ensemble models to improve landslide susceptibility mapping reliability. Bull. Eng. Geol. Environ., 82, 309. DOI: 10.1007/s10064-023-03328-8

Turker, E.: 2025, Prediction of the safety factor and the failure surface for slope stability analysis using Jaya optimization method. Geomech. Eng., 40, 339. DOI: 10.12989/gae.2025.40.5.339

Ullah, M., Qiu, H., Huangfu, W., Yang, D., Wei, Y. and Tang, B.: 2025, Integrated machine learning approaches for landslide susceptibility mapping along the Pakistan–China Karakoram Highway. Land, 14, 172. DOI: 10.3390/land14010172

UNDRR: 2022, Annual Report 2022. Geneva, Switzerland.

Van Tien, P., Luong, L.H., Nhat, L.M., Thanh, N.K. and Van Cuong, P.: 2021, Landslides along Halong-Vandon Expressway in Quang Ninh Province, Vietnam. In: Guzzetti, F., Mihalić Arbanas, S., Reichenbach, P., Sassa, K., Bobrowsky, P.T. and Takara, K. (Eds.), Understanding and reducing landslide disaster risk, 2. Springer, Cham, 133–139. DOI: 10.1007/978-3-030-60227-7_14

Wadadar, S. and Mukhopadhyay, B.P.: 2022, GIS-based landslide susceptibility zonation and comparative analysis using analytical hierarchy process and conventional weighting-based multivariate statistical methods in the Lachung River Basin, North Sikkim. Nat. Hazards, 113, 1199–1236. DOI: 10.1007/s11069-022-05344-5

Ward, P.J., Blauhut, V., Bloemendaal, N., Daniell, J.E., de Ruiter, M.C., Duncan, M.J., Emberson, R., Jenkins, S.F., Kirschbaum, D., Kunz, M., Mohr, S., Muis, S., Riddell, G.A., Schäfer, A., Stanley, T., Veldkamp, T.I.E. and Winsemius, H.C.: 2020, Natural hazard risk assessments at the global scale. Nat. Hazards Earth Syst. Sci., 20, 1069–1096. DOI: 10.5194/nhess-20-1069-2020

Wardhani, N.W.S., Rochayani, M.Y., Iriany, A., Sulistyono, A.D. and Lestantyo, P.: 2019, Cross-validation metrics for evaluating classification performance on imbalanced data. 2019 International Conference on Computer, Control, Informatics and its Applications (IC3INA), 14–18. DOI: 10.1109/IC3INA48034.2019.8949568

Widiastuti, M., Widodo, S., Tumpu, M., Hatta, M.P. and Alimuddin, I.: 2025, Landslide hazard index assessment using GIS and AHP in Tapalang, Mamuju, West Sulawesi Province, Indonesia. Eng. Technol. Appl. Sci. Res., 15, 24857–24863. DOI: 10.48084/etasr.10835

Yadav, M., Pal, S.K., Singh, P.K. and Gupta, N.: 2023, Landslide susceptibility zonation mapping using frequency ratio, information value model, and logistic regression model: A case study of Kohima District in Nagaland, India. In: Thambidurai, P. and Singh, T.N. (Eds.), Landslides: detection, prediction and monitoring: Technological Developments. Springer, Cham, 333–363. DOI: 10.1007/978-3-031-23859-8_17

Yang, C., Liu, L.-L., Huang, F., Huang, L. and Wang, X.-M.: 2023, Machine learning-based landslide susceptibility assessment with optimized ratio of landslide to non-landslide samples. Gondwana Res., 123, 198–216. DOI: 10.1016/j.gr.2022.05.012

Yilmaz, I.: 2009, Landslide susceptibility mapping using frequency ratio, logistic regression, artificial neural networks and their comparison: A case study from Kat landslides (Tokat—Turkey). Comput. Geosci., 35, 1125–1138. DOI: 10.1016/j.cageo.2008.08.007

Yu, L., Wang, Y. and Pradhan, B.: 2024, Enhancing landslide susceptibility mapping incorporating landslide typology via stacking ensemble machine learning in Three Gorges Reservoir, China. Geosci. Front., 15, 101802. DOI: 10.1016/j.gsf.2024.101802

Zeng, T., Wu, L., Peduto, D., Glade, T., Hayakawa, Y.S. and Yin, K.: 2023, Ensemble learning framework for landslide susceptibility mapping: Different basic classifier and ensemble strategy. Geosci. Front., 14, 101645. DOI: 10.1016/j.gsf.2023.101645

Zhang, K., Wang, S., Bao, H. and Zhao, X.: 2019, Characteristics and influencing factors of rainfall-induced landslide and debris flow hazards in Shaanxi Province, China. Nat. Hazards Earth Syst. Sci., 19, 93–105. DOI: 10.5194/nhess-19-93-2019

Zhang, Q., Ning, Z., Ding, X., Wu, J., Wang, Z., Tsangaratos, P., Ilia, I., Wang, Y. and Chen, W.: 2024, Hybrid integration of bagging and decision tree algorithms for landslide susceptibility mapping. Water, 16, 657. DOI: 10.3390/w16050657

Zhao, F., Miao, F., Wu, Y., Ke, C., Gong, S. and Ding, Y.: 2024, Refined landslide susceptibility mapping in township area using ensemble machine learning method under dataset replenishment strategy. Gondwana Res., 131, 20–37. DOI: 10.1016/j.gr.2024.02.011

Zhou, Z., Duan, J., Geng, S. and Li, R.: 2024, The role of highway construction in influencing agricultural green total factor productivity in China: agricultural industry structure transformation perspective. Front. Sustain. Food Syst., 7. DOI: 10 .3389/fsufs.2023.1315201