Journal of Science and Transport Technology
Journal homepage: https://jstt.vn/index.php/en

UNIVERSITY OF TRANSPORT TECHNOLOGY

Article info
Type of article:
Original research paper

DOI:
https://doi.org/10.58845/jstt.utt.2026.en.6.2.455-473

*__Corresponding author:__
Email address:
tphong1617@gmail.com

Received: 15/12/2025
Received in Revised Form: 11/05/2026
Accepted: 18/06/2026

CatBoost-Based Landslide Susceptibility Modeling Using Landsat 8 Imagery: A Case Study from Phuoc Son, Vietnam

Tran Huu Anh[1], Tran Van Phong[2,3*], Dong Doan Van[4], Phan Trong Trinh[2,3], Nguyen Thi Bich Hanh[5], Indra Prakash[6]

[1]Faculty of Technology and Engineering, Thai Binh University, Hung Yen, Vietnam; anhuni1006@gmail.com

[2]Institute of Earth Sciences, Vietnam Academy of Science and Technology, 68 Huynh Thuc Khang, Hanoi, Vietnam; phantrongt@yahoo.com

[3]Graduate University of Science and Technology, Vietnam Academy of Science and Technology, 18 Hoang Quoc Viet, Hanoi, Vietnam

[4]University of Transport Ho Chi Minh City, 2 Vo Oanh, Thanh My Tay, Ho Chi Minh 700000, Vietnam; dongdv@ut.edu.vn

[5]Geotechnical Engineering and Artificial Intelligence research Group, University of Transport Technology, Hanoi, Vietnam; hanhntb@utt.edu.vn

[6]Formerly Dy. Director General, Geological Survey of India, Gujarat, India; indra52prakash@gmail.com

Abstract: The mountainous terrain and monsoon-dominated climate of Central Vietnam make the region highly susceptible to rainfall-induced landslides, particularly in the Phuoc Son area, where steep slopes, complex geological conditions, and intense precipitation frequently trigger slope failures. This study develops a landslide susceptibility model using the CatBoost machine learning algorithm by integrating Landsat 8-derived surface indicators with key geo-environmental factors, including topographic, geological, hydrological, and vegetation-related parameters such as the Normalized Difference Vegetation Index (NDVI). The predictive performance of the proposed model was evaluated and compared with four widely used approaches, namely Logistic Regression (LR), Support Vector Machine (SVM), Multilayer Perceptron (MLP), and Deep Neural Network (DNN). Model accuracy was assessed using the Area Under the Receiver Operating Characteristic Curve (AUC). The results demonstrated that CatBoost outperformed all benchmark models, achieving AUCs of 0.97 and 0.93 on the training and testing datasets, respectively, indicating excellent predictive capability and strong generalization. The resulting landslide susceptibility maps effectively delineated areas with varying levels of landslide risk and provided enhanced spatial accuracy in identifying highly susceptible zones. These findings highlight the effectiveness of integrating remote sensing data with advanced ensemble machine learning techniques for landslide susceptibility assessment in tropical mountainous environments and provide a reliable scientific basis for hazard mitigation, land-use planning, and disaster risk management in Central Vietnam.

Keywords: CatBoost, Landsat 8, Remote Sensing, Phuoc Son, Landslide Susceptibility, Vietnam.

1. Introduction

Landslides are among the most damaging geo-hazards in tropical mountainous regions, where steep terrain, intense monsoon rainfall, and rapid land-use changes collectively destabilize slopes [1-8]. Central Vietnam is particularly vulnerable, as frequent typhoons and extreme precipitation events trigger widespread slope failures, posing significant risks to infrastructure, hydropower development, and human settlements [3, 9-11]. In this context, landslide susceptibility mapping, which identifies areas prone to future slope failures, has become a critical tool for hazard mitigation, land-use planning, and disaster risk management [3, 5, 6, 11-13].

The development of reliable susceptibility models depends on the accurate representation of landslide conditioning factors, including topographic, geological, hydrological, and land-cover variables, along with a comprehensive landslide inventory [14-16]. Traditional approaches based on field surveys and expert interpretation provide valuable local insights but are often time-consuming, costly, and difficult to apply consistently over large or inaccessible areas [8, 17, 18]. The integration of remote sensing and GIS techniques has significantly improved this process by enabling the efficient extraction of terrain attributes and surface indicators from satellite data [19-23]. In particular, Landsat 8 imagery offers valuable multispectral information for deriving vegetation and surface condition indices (e.g., NDVI), allowing consistent and repeatable regional-scale analysis [23, 24].

Conventional susceptibility modeling approaches, including qualitative methods (e.g., AHP and fuzzy logic) and statistical techniques such as logistic regression, have been widely used but often struggle to capture the complex and nonlinear relationships governing landslide occurrence [25]. In recent years, machine learning (ML) methods, including support vector machines (SVM), multilayer perceptrons (MLPs), and deep neural networks (DNN), have gained prominence due to their ability to model nonlinear interactions among conditioning factors [3, 5]. However, these models may still face challenges related to overfitting, parameter tuning, and handling heterogeneous datasets [18, 26, 27].

Among advanced ML approaches, ensemble learning techniques, particularly gradient-boosting algorithms, have demonstrated superior predictive performance in landslide susceptibility assessment [28-30]. CatBoost, a modern boosting algorithm, is specifically designed to handle categorical and heterogeneous data efficiently while reducing prediction bias and overfitting [31-33]. Despite its demonstrated effectiveness in various environmental modeling applications, its integration with Landsat 8–derived surface indicators for landslide susceptibility mapping in tropical monsoon environments such as Central Vietnam remains limited [34-36].

Therefore, this study aims to develop an integrated landslide susceptibility modeling framework by combining Landsat 8–derived surface indicators with the CatBoost algorithm for the Phuoc Son region. The performance of the proposed model is evaluated against conventional methods, including logistic regression, SVM, MLP, and DNN. The novelty of this work lies in the integration of multispectral remote sensing–based feature extraction with an advanced ensemble learning approach to improve predictive accuracy and spatial reliability in a data-scarce, high-risk mountainous environment.

2. Study area

The study area is located in the Phuoc Son region, currently administratively affiliated with Da Nang City in Central Vietnam (formerly Phuoc Son District, Quang Nam Province), within a mountainous area characterized by complex terrain and high landslide susceptibility (Fig. 1). The region is situated in the tropical monsoon climatic zone and experiences intense seasonal rainfall associated with typhoons and prolonged precipitation events, which are recognized as the primary triggers of slope instability [3].

Topographically, the region is highly rugged, consisting of steep slopes, deeply incised valleys, and narrow ridges. Elevations range from low-lying valley floors of a few tens of meters above sea level to mountainous peaks exceeding 1,500 m, with some summits rising above 2,000 m. Slope gradients vary significantly, from nearly flat areas in valley bottoms to very steep slopes (>40°), contributing to widespread instability across the landscape.

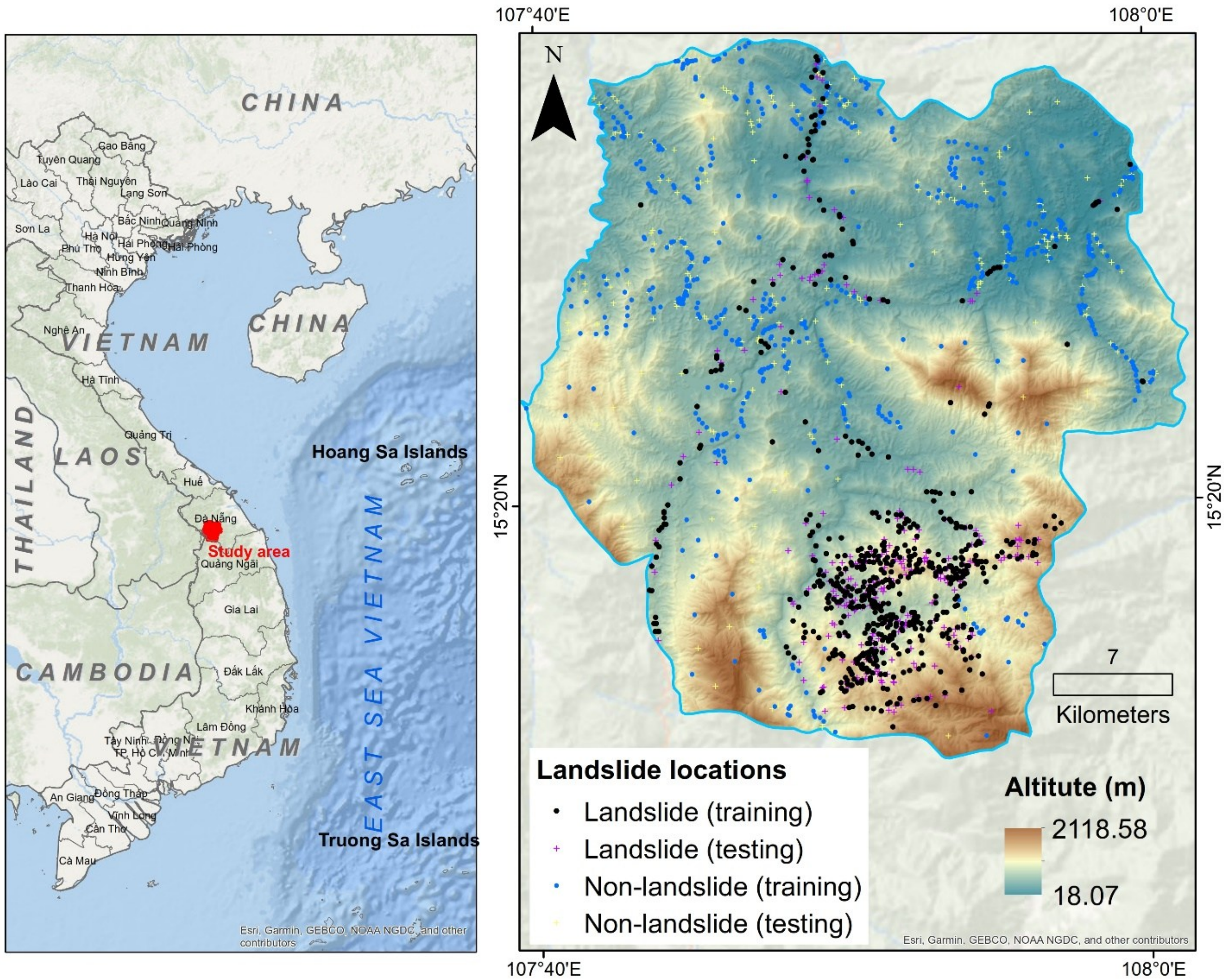

Fig. 1. Research area map showing sample locations of Landslide and non-landslide points

The geological setting of Phuoc Son is diverse, comprising a mixture of hard bedrock formations and weathered, structurally weakened materials that are highly susceptible to failure under intense rainfall conditions. Variations in lithology, combined with structural discontinuities and weathering profiles, play a critical role in controlling landslide occurrence.

Land cover in the region is heterogeneous, dominated by dense tropical forests in the uplands, with localized areas of agricultural land, shrubland, and exposed surfaces in lower elevations and inhabited zones. Vegetation conditions, represented by indices such as NDVI, vary from dense canopy cover to sparsely vegetated areas, influencing slope stability through root reinforcement and hydrological regulation.

Overall, the combination of steep terrain, complex geology, intense rainfall, and heterogeneous land cover makes Phuoc Son a highly landslide-prone region and an appropriate case study for evaluating advanced susceptibility modeling approaches.

3. Methodology and Materials

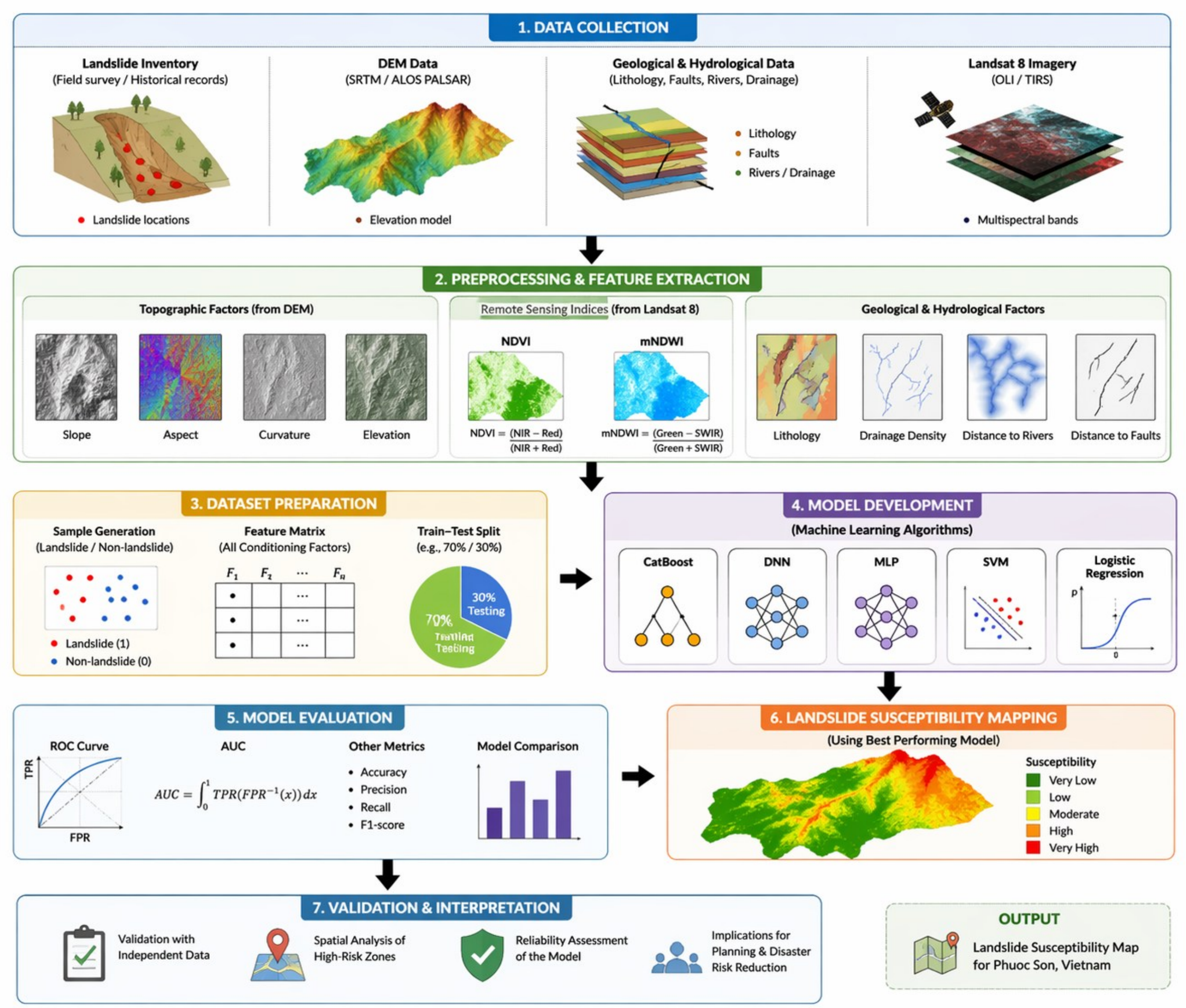

Fig. 2. Workflow of landslide susceptibility modeling using Landsat 8 data and machine learning models

This study develops an integrated landslide susceptibility modeling framework by combining Landsat 8–derived surface indicators with advanced machine learning techniques, with particular emphasis on the CatBoost algorithm (Fig. 2). The methodology involves (i) preparation of a landslide inventory and conditioning factors, (ii) extraction of remote sensing–based indices, (iii) model development using CatBoost and benchmark algorithms, and (iv) performance evaluation using standard accuracy metrics. The overall workflow is designed to assess the effectiveness of CatBoost in capturing complex nonlinear relationships governing landslide occurrence in a tropical mountainous environment.

3.1. Machine Learning Models

3.1.1. CatBoost Model

CatBoost is a gradient boosting decision-tree algorithm designed to efficiently handle heterogeneous datasets containing both numerical and categorical variables [33, 37]. It builds an ensemble of decision trees sequentially, where each tree minimizes the residual errors of the previous model, thereby improving predictive accuracy. The general formulation of gradient boosting is expressed as:

$$F_m(x) = F_{m-1}(x) + \eta \cdot h_m(x) \quad (1)$$

where:

$F_m(x)$= model prediction at iteration m,
$h_m(x)$= decision tree at iteration m,
η= learning rate.

At each iteration, the model minimizes a loss function, $L(y, F(x))$ by computing pseudo-residuals:

$$r_i^{(m)} = -\frac{\partial L(y_i, F(x_i))}{\partial F(x_i)} \quad (2)$$

These residuals are used to train subsequent trees, enabling the model to capture complex nonlinear relationships among conditioning factors.

In this study, CatBoost integrates topographic, geological, hydrological, and remote sensing variables to predict landslide susceptibility. Its ability to handle categorical and continuous variables without extensive preprocessing makes it particularly suitable for geospatial applications.

A key advantage of CatBoost is its ordered boosting mechanism, which reduces overfitting and prediction bias by preventing information leakage during training. Additionally, it employs symmetric (oblivious) trees, which enhance computational efficiency and generalization.

In this study, CatBoost was used as the primary model to predict landslide susceptibility by integrating multi-source geospatial data, including topographic, geological, hydrological, and remote sensing–derived variables. Landsat 8 imagery was utilized to derive vegetation and moisture indices such as NDVI and mNDWI, which are important indicators of slope stability.

The model was trained using landslide and non-landslide samples, allowing it to learn complex nonlinear interactions among conditioning factors. Its ability to handle mixed data types and capture high-order relationships makes it particularly suitable for landslide susceptibility assessment.

3.1.2. Benchmark Models

To evaluate the performance of CatBoost, four widely used machine learning models were implemented for comparison:

Deep Neural Network (DNN): A multilayer network capable of modeling complex nonlinear relationships through hierarchical feature learning. It processes multiple conditioning factors simultaneously to estimate landslide probability [38].

Multilayer Perceptron (MLP): A feedforward neural network with hidden layers that captures nonlinear interactions between input variables and landslide occurrence [39].

Support Vector Machine (SVM): A supervised learning model that uses kernel functions to map input data into higher-dimensional space and identify optimal decision boundaries for classification [40].

Logistic Regression (LR): A statistical baseline model that estimates landslide probability based on a linear relationship between predictors and the log-odds of occurrence [41].

These models were selected due to their widespread application in landslide susceptibility studies and their ability to represent different levels of model complexity.

Among these, logistic regression provides a statistical baseline, while DNN and MLP capture nonlinear relationships, and SVM enables kernel-based classification.

The logistic regression model is expressed as:

$$(y = 1 \mid x) = \frac{1}{1 + e^{-(\beta_0 + \sum_{i=1}^{n} \beta_i x_i)}} \quad (3)$$

where:

$P(y = 1 \mid x)$= probability of landslide occurrence,
β_i= regression coefficients,
x_i= conditioning factors.

3.2. Landslide Conditioning Factors

The model input variables include topographic, geological, hydrological, and land-cover factors derived from GIS and remote sensing data. Topographic parameters such as slope, aspect, elevation, and curvature were extracted from DEM data, while geological and hydrological variables included lithology, distance to faults, drainage density, and proximity to rivers.

Landsat 8 imagery was used to derive key surface indicators, including:

Normalized Difference Vegetation Index (NDVI): representing vegetation density

$$\mathrm{NDVI} = \frac{\mathrm{NIR} - \mathrm{Red}}{\mathrm{NIR} + \mathrm{Red}} \tag{4}$$

where NIR and Red represent near-infrared and red spectral bands, respectively.

Modified Normalized Difference Water Index (mNDWI): indicating surface moisture conditions

$$\mathrm{mNDWI} = \frac{\mathrm{Green} - \mathrm{SWIR}}{\mathrm{Green} + \mathrm{SWIR}} \tag{5}$$

where Green and SWIR represent the green and short-wave infrared bands.

These indices provide important information on vegetation cover and soil moisture, both of which significantly influence slope stability.

3.3. Model Training and Validation

The dataset was divided into training and testing subsets (70/30) to evaluate model performance and generalization capability. The models were trained using landslide occurrence data and corresponding conditioning factors, and their predictive performance was assessed using standard evaluation metrics, including the Area Under the ROC Curve (AUC) and overall accuracy.

The AUC is defined as:

$$\mathrm{AUC} = \int_0^1 \mathrm{TPR}(\mathrm{FP\,R}^{-1}(\mathrm{x}))\,\mathrm{dx} \tag{6}$$

where:

TPR = True Positive Rate,

FPR = False Positive Rate.

3.4. Methodological Advantages

The proposed CatBoost-based framework offers several advantages, including efficient handling of heterogeneous data, strong capability to model nonlinear relationships, reduced overfitting through ordered boosting, and improved predictive accuracy compared to conventional models.

4. Results and analysis

4.1. Conditioning Factors and Their Spatial Characteristics

A total of twelve geo-environmental conditioning factors were compiled to characterize landslide susceptibility in the Phuoc Son region (Fig. 3). These include topographic, geological, hydrological, and anthropogenic variables derived from GIS and remote sensing datasets.

Rainfall ranges from approximately 3125 mm to 4046 mm annually, with higher precipitation zones showing strong correspondence with landslide occurrences. Elevation varies from 18 m to over 2100 m, reflecting rugged terrain conditions, particularly in the southern mountainous region, where landslide density is higher. Slope angle is a dominant factor, with most landslides occurring on slopes exceeding 30°, confirming the role of gravitational instability.

Geological and soil-related parameters significantly influence slope stability. Landslides are concentrated in weathered formations such as sialferite and unconsolidated sediments, as well as clay-rich soils with low shear strength. Similarly, proximity-based factors indicate that landslides frequently occur within 200 m of faults and 100 m of roads, highlighting the combined influence of tectonic activity and anthropogenic disturbances.

Terrain curvature and aspect also contribute to landslide susceptibility. Concave slopes favor water accumulation, increasing instability, while slope orientation influences moisture retention and vegetation growth. Land-use patterns derived from Landsat 8 reveal that disturbed areas, including agricultural land and built-up zones, exhibit higher landslide susceptibility compared to densely forested regions.

Overall, these conditioning factors collectively capture the complex interactions governing slope instability and form a robust basis for machine learning modeling.

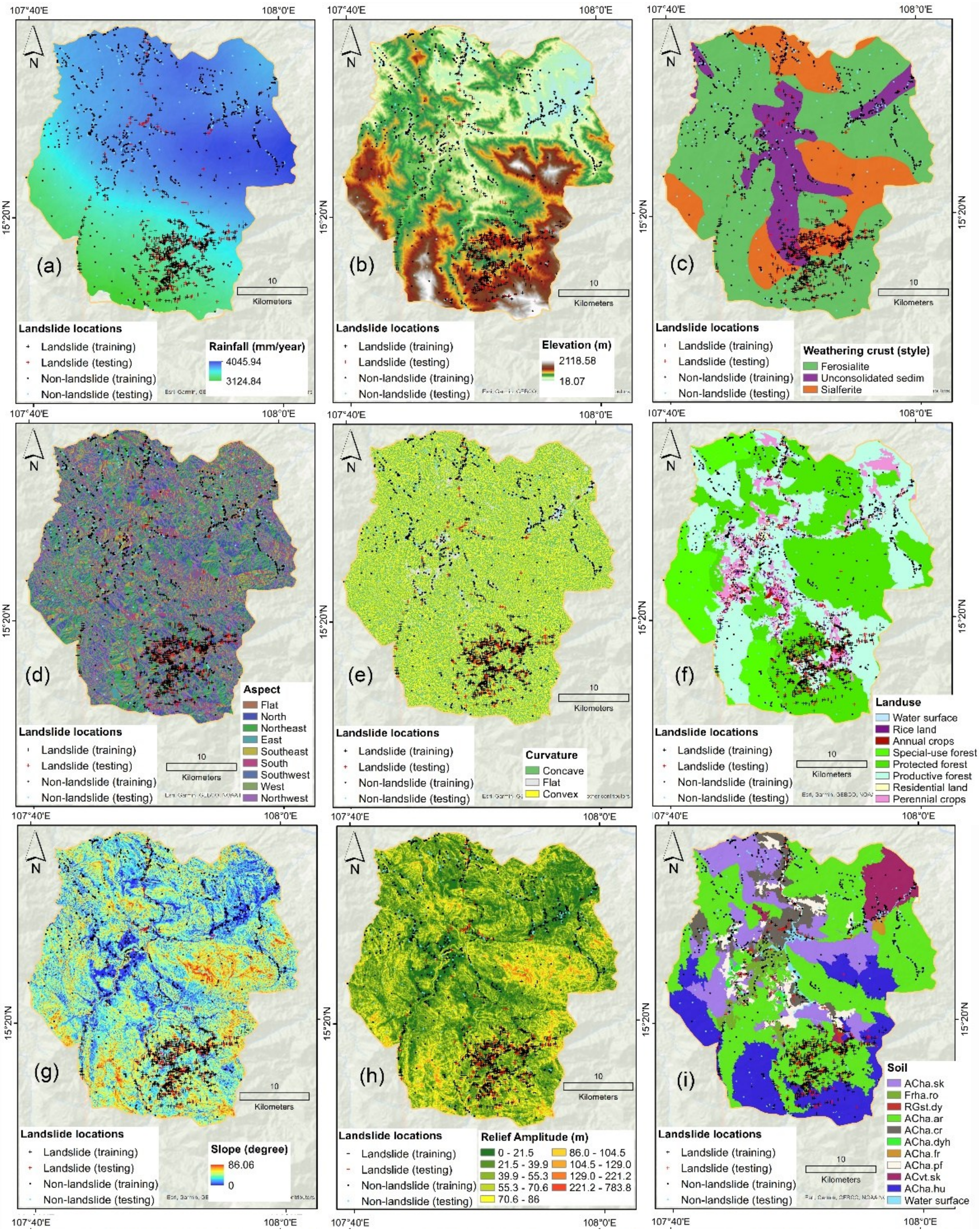

Fig. 3. Spatial distribution of landslide conditioning factors in the Phuoc Son region

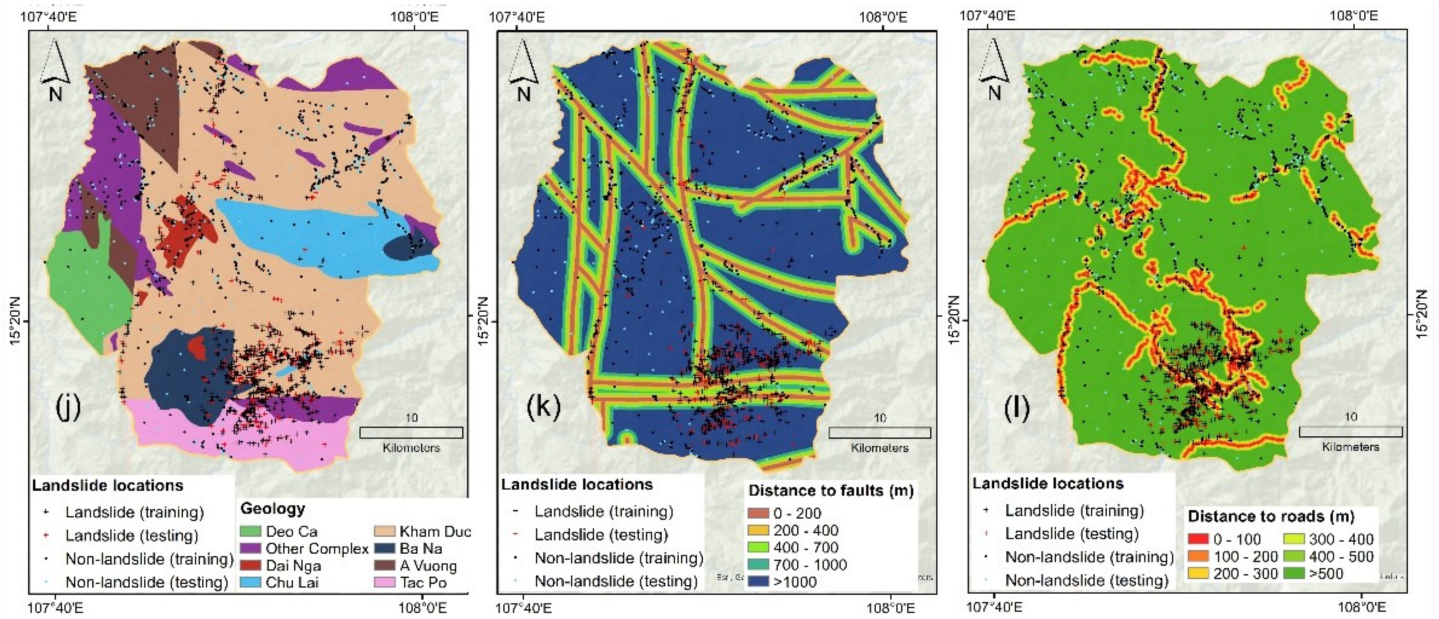

Fig. 3. (continued)

Collectively, seven Landsat 8 bands represent a diverse and complementary spectral dataset for modeling landslide susceptibility. Their variation across the study area, as illustrated in the maps, provides critical input for machine learning models such as Logistic Regression or CatBoost, enabling the capture of surface conditions and geomorphological anomalies associated with landslide occurrence. The spatial overlap of reflectance anomalies with known landslide locations supports their relevance as predictive features.

4.2. Landsat 8 Spectral Characteristics

The seven multispectral bands of Landsat 8 imagery provide critical information on surface conditions influencing landslide occurrence (Fig. 4). Each band captures distinct spectral properties related to vegetation, soil moisture, and land cover.

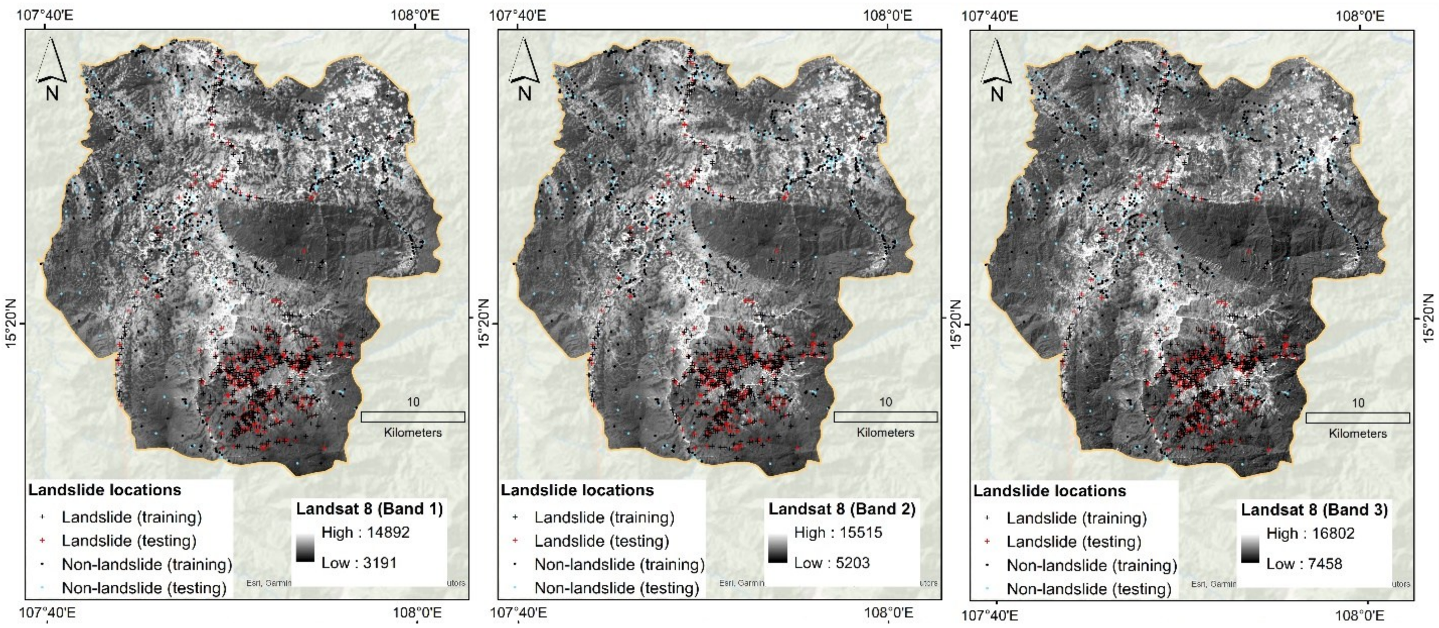

Fig. 4. Landsat 8 multispectral bands were used for landslide susceptibility modeling in the Phuoc Son region

Lower reflectance values in visible and near-infrared bands correspond to dense vegetation and stable slopes, whereas higher reflectance values indicate exposed soil, disturbed land, or degraded vegetation conditions commonly associated with landslide-prone areas. Shortwave infrared bands (SWIR1 and SWIR2) highlight variations in soil moisture and mineral composition, which are key indicators of slope instability.

The integration of these spectral features enhances the model's ability to detect subtle environmental changes associated with landslide processes, particularly vegetation degradation and moisture anomalies.

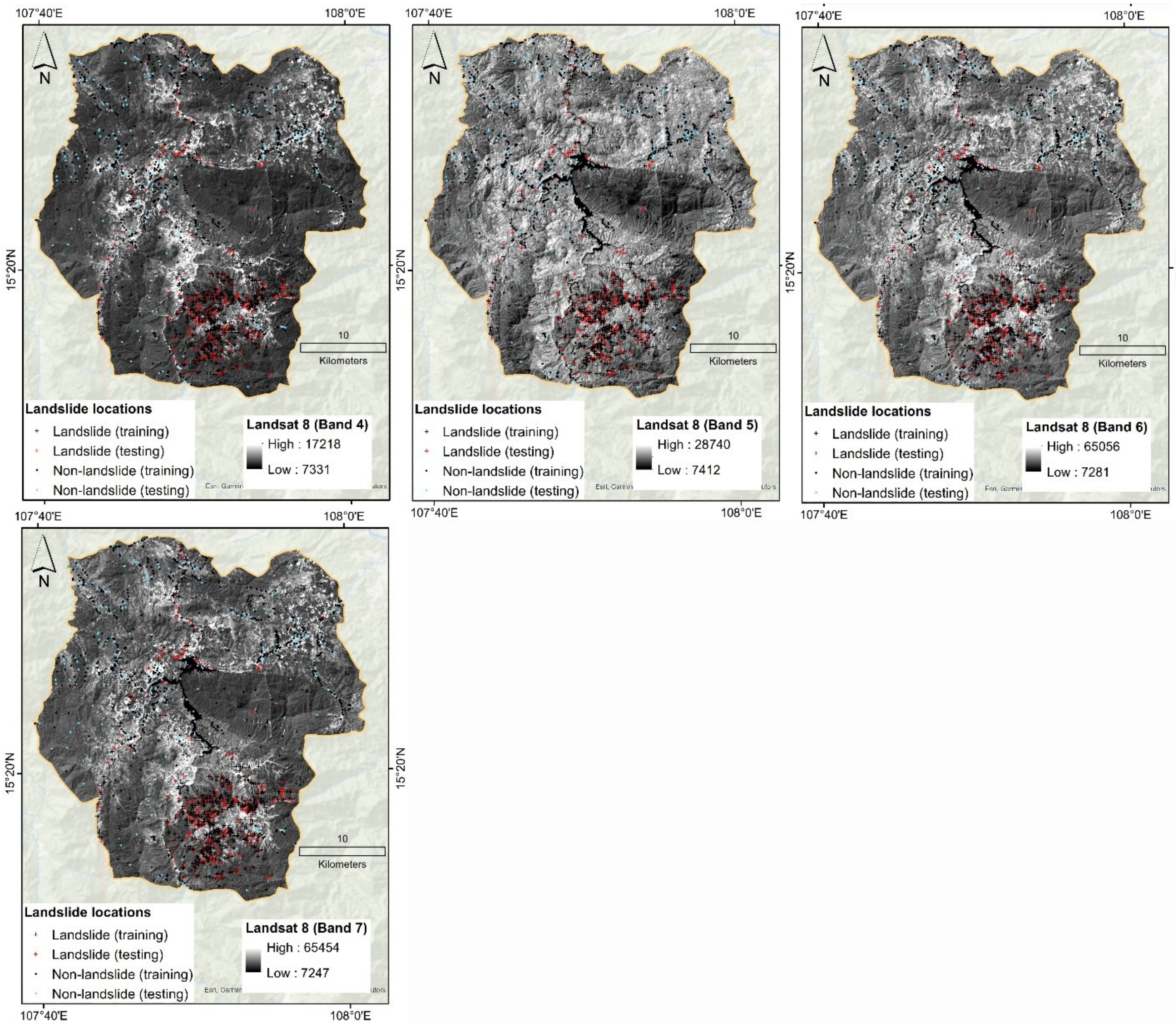

Fig. 4. (continued)

4.3. Model Performance Based on ROC Analysis

The comparative analysis of ROC curves reveals clear differences in model performance and generalization capability (Fig. 5). Among the five models evaluated, CatBoost achieved the highest AUC values on both training (0.97) and testing (0.93) datasets, indicating superior predictive accuracy and robustness.

SVM also demonstrated strong performance with high and stable AUC values, reflecting good generalization. In contrast, logistic regression exhibited the lowest AUC, highlighting its limitation in capturing nonlinear relationships. Neural network models (MLP and DNN) showed improved performance over LR but displayed noticeable differences between training and testing results, suggesting moderate overfitting.

Among the five models (SVM, LR, MLP, DNN, CatBoost), CatBoost achieved the best balance between accuracy and generalization, maintaining high sensitivity and specificity across the full range of false positive rates.

4.4. Multi-Metric Model Evaluation

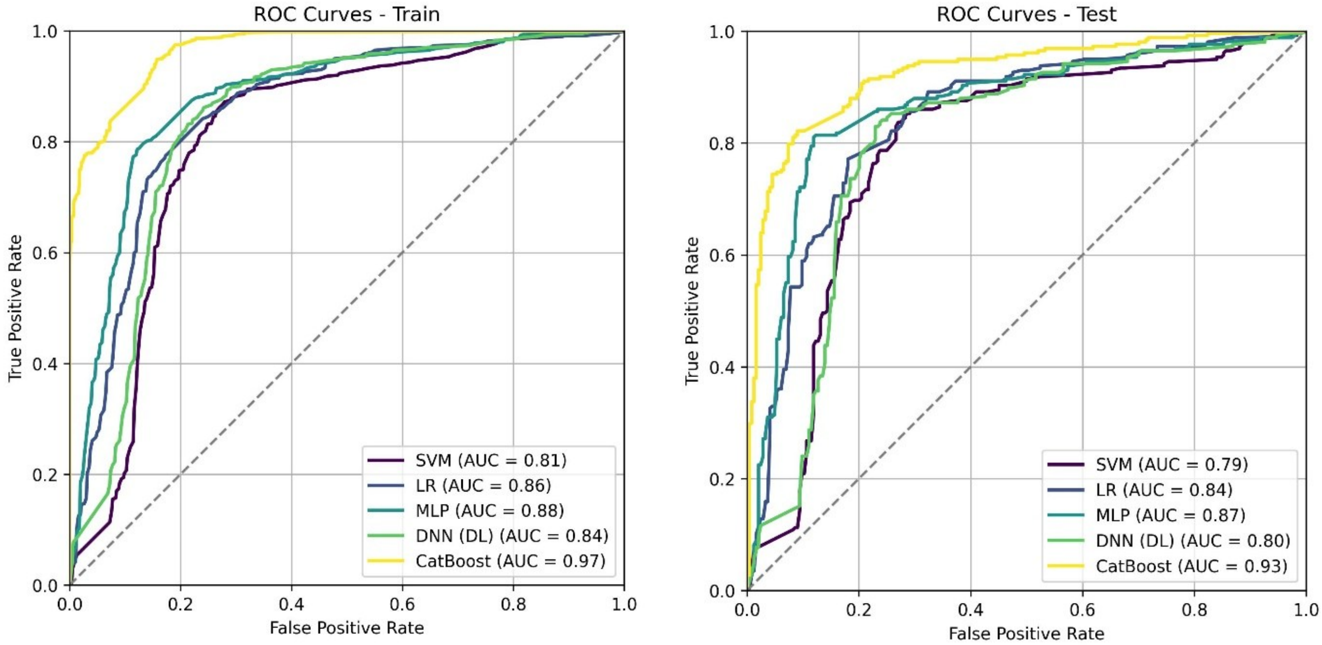

Fig. 5. ROC curves comparing model performance for landslide susceptibility prediction

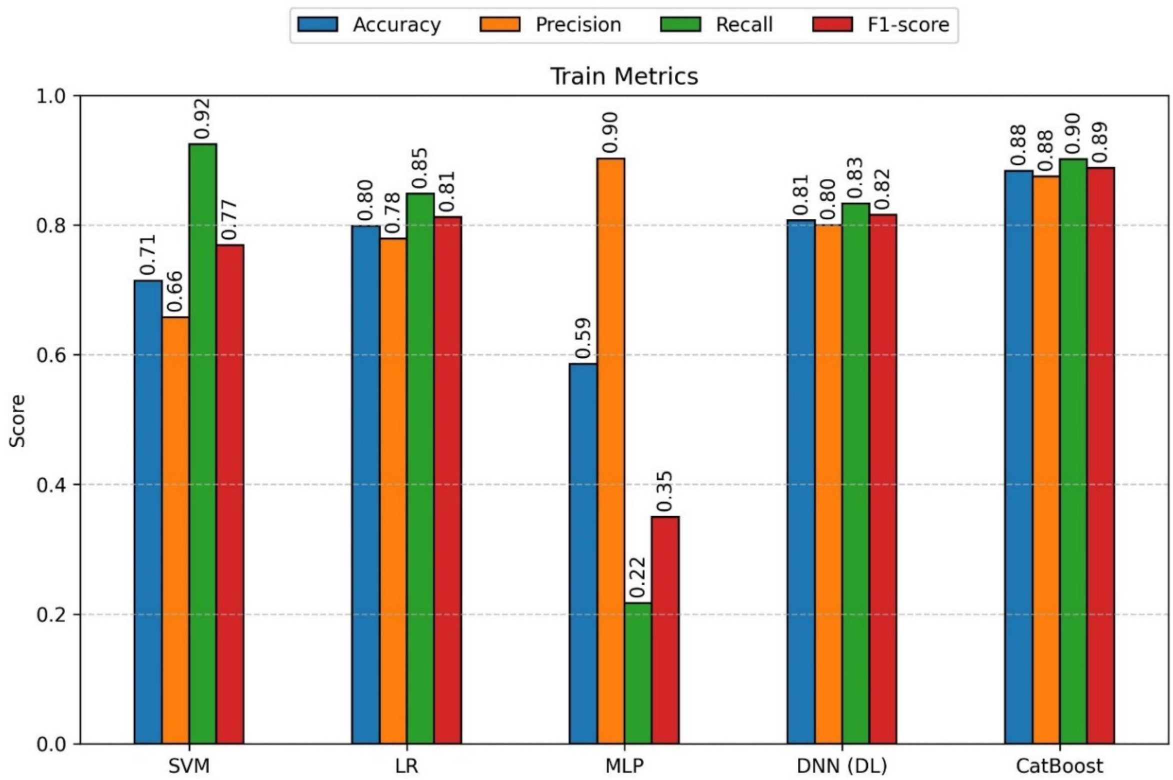

Fig. 6. Comparative performance of machine learning models based on multiple evaluation metrics

The evaluation using multiple performance metrics (Accuracy, Precision, Recall, F1-score, Kappa, MCC, RMSE, Jaccard Index, and Brier Score) further confirms the superiority of the CatBoost model (Fig. 6).

Logistic regression produced consistent but lower performance across all metrics, indicating stable yet limited predictive capability. SVM and MLP achieved improved results, with SVM showing better generalization and MLP exhibiting slight overfitting.

The DNN model achieved the highest training performance but showed a noticeable drop in testing accuracy, reflecting overfitting due to its high model complexity. In contrast, CatBoost consistently achieved the highest testing accuracy and lowest error values, demonstrating both strong predictive capability and excellent generalization.

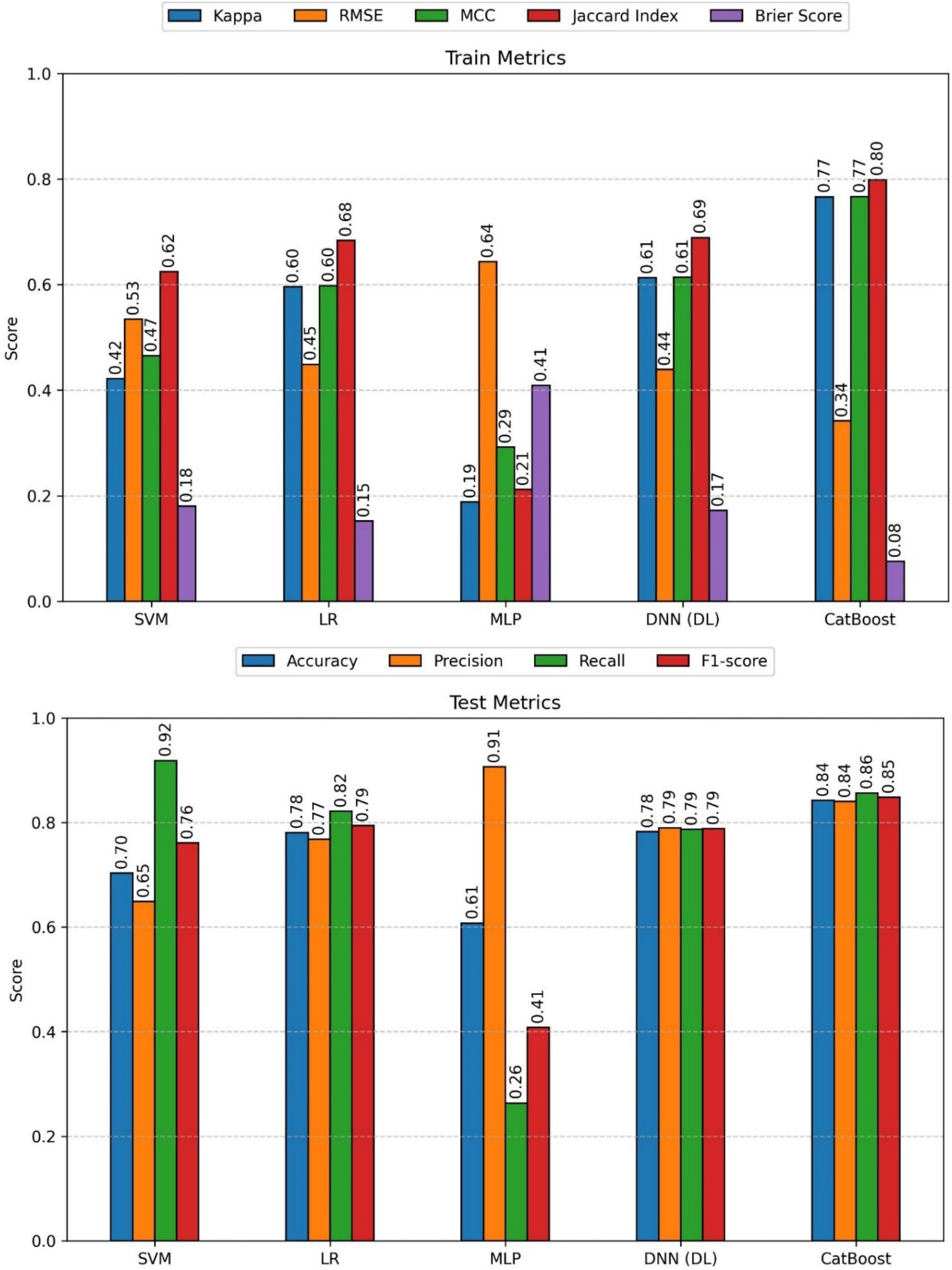

Fig. 6. (continued)

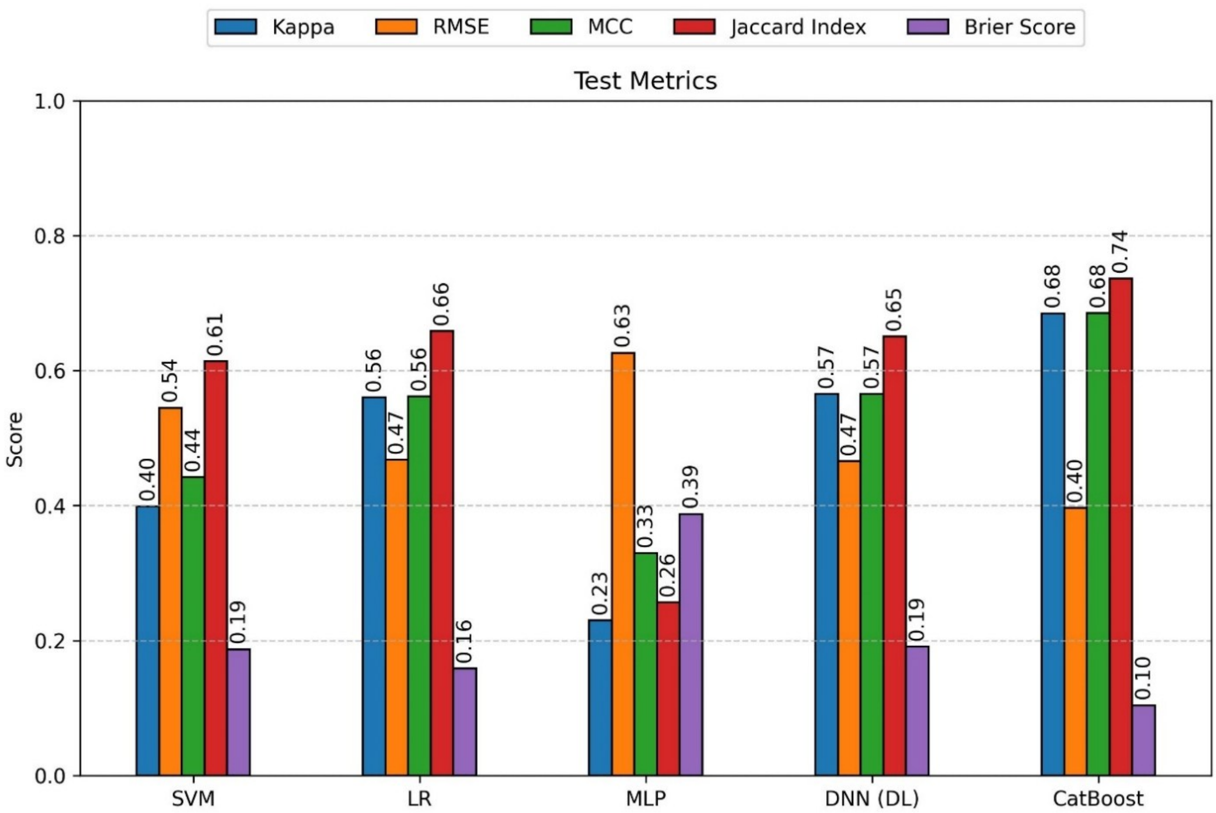

Fig. 6. (continued)

These results highlight that ensemble learning methods provide a more reliable and balanced approach for landslide susceptibility modeling compared to conventional statistical and deep learning methods.

4.5. Landslide Susceptibility Mapping

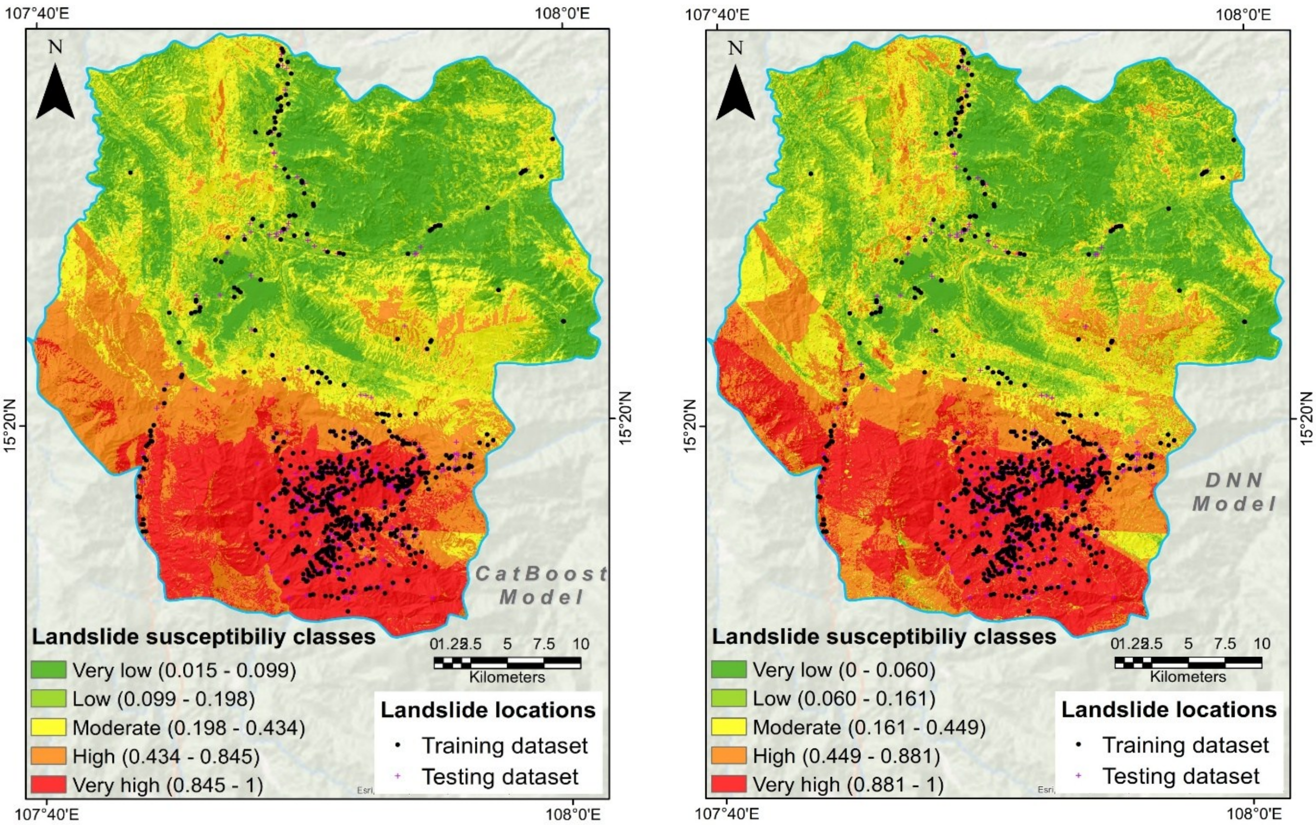

Fig. 7. Landslide susceptibility maps generated using different machine learning models

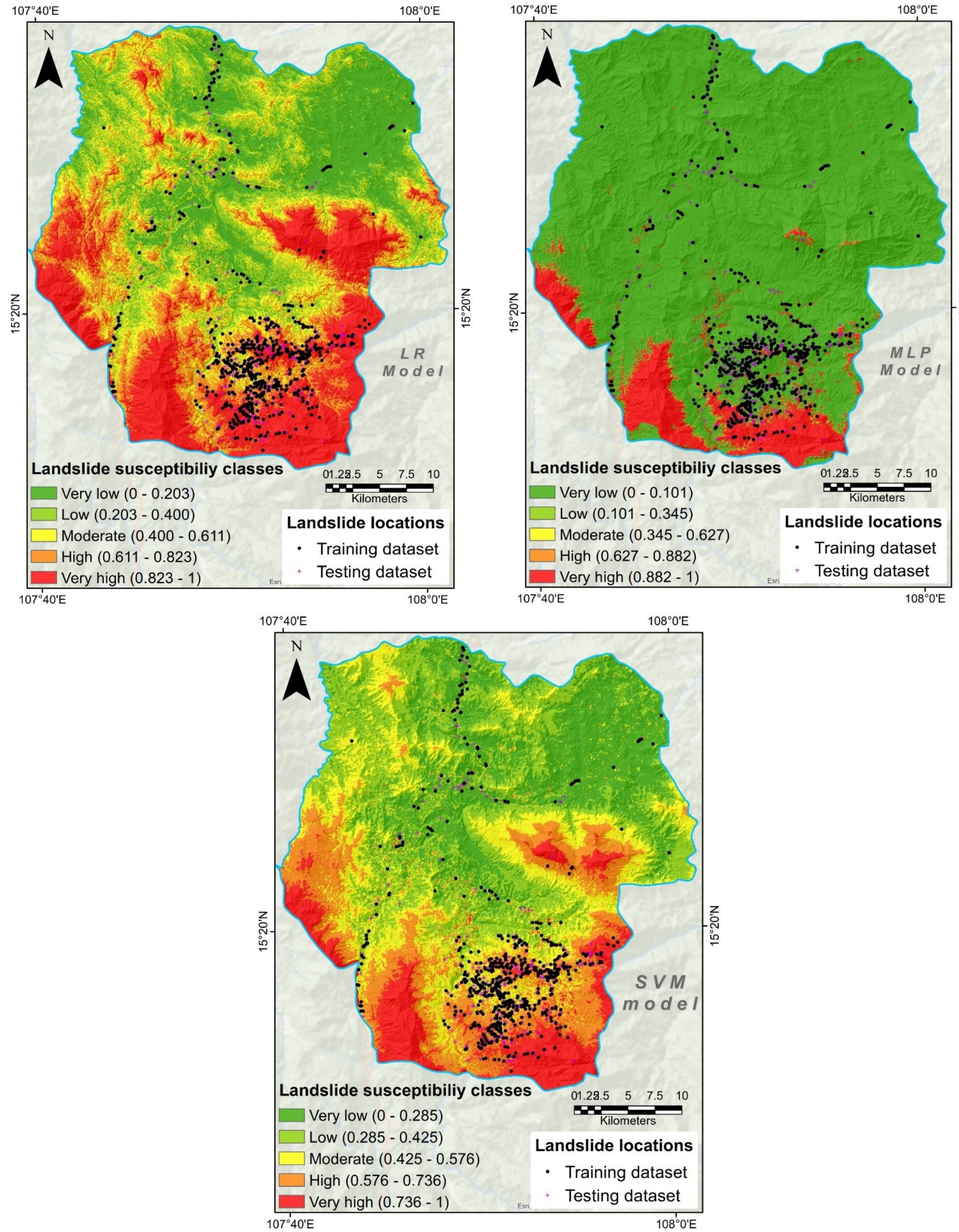

Fig. 7. (continued)

The susceptibility maps generated by different models show both consistent patterns and notable differences (Fig. 7). All models identify steep slopes, sparsely vegetated areas, and disturbed land as high-risk zones, confirming the dominant role of slope, vegetation, and land cover.

However, spatial resolution and prediction detail vary significantly among models. Logistic regression produces smoother and more generalized susceptibility patterns, while CatBoost and DNN generate more detailed and spatially refined predictions. CatBoost, in particular, captures localized high-risk zones with greater precision, closely matching observed landslide locations.

SVM and MLP provide intermediate results, balancing generalization and detail, but not achieving the same level of spatial accuracy as CatBoost. The improved performance of CatBoost is attributed to its ability to model nonlinear interactions among conditioning factors.

Overall, the susceptibility maps confirm that advanced machine learning models, particularly CatBoost, provide more accurate and reliable delineation of landslide-prone areas.

5. Discussion

The results demonstrate that the integrated CatBoost–Landsat 8 framework provides a robust and highly accurate approach for landslide susceptibility mapping in the Phuoc Son region. The model achieved excellent predictive performance (AUC = 0.97 for training and 0.93 for testing), significantly outperforming conventional statistical approaches such as logistic regression and showing competitive or superior performance relative to advanced machine learning models [3]. This improvement can be attributed to the ability of CatBoost to effectively capture nonlinear relationships and complex interactions among conditioning factors, which are characteristic of landslide processes in tropical mountainous environments [42].

A key observation is the strong agreement between predicted susceptibility patterns and the spatial distribution of known landslides. High-susceptibility zones are consistently identified along steep slopes, deeply incised valleys, and areas affected by anthropogenic disturbances such as road construction and deforestation [22]. This spatial consistency suggests that the model successfully represents the dominant geomorphological and environmental controls governing slope instability. Unlike simpler models that produce generalized susceptibility patterns, CatBoost provides more spatially refined predictions, allowing the identification of localized high-risk zones that are critical for hazard mitigation.

The comparative analysis further highlights differences in model behavior. Logistic regression exhibited stable but limited predictive capability due to its linear assumptions, while SVM and MLP provided improved performance by capturing nonlinear patterns [43]. The DNN achieved high training accuracy but showed a noticeable generalization gap, indicating sensitivity to overfitting. In contrast, CatBoost demonstrated both high accuracy and strong generalization, reflecting the effectiveness of boosting-based ensemble learning in balancing model complexity and robustness. These findings align with recent studies indicating that tree-based ensemble methods often outperform both traditional statistical models and deep learning approaches in geospatial hazard prediction tasks.

The integration of Landsat 8–derived spectral indices represents a significant contribution of this study. Vegetation and moisture-related indicators, particularly NDVI and NDWI, were found to play a critical role in improving model performance [44]. Low NDVI values, associated with sparse or degraded vegetation, correspond to reduced root cohesion and increased susceptibility to surface erosion and slope failure. Similarly, high moisture conditions inferred from NDWI are indicative of elevated pore-water pressure, which reduces shear strength and promotes landslide initiation. The inclusion of these indices enables the model to account for surface and environmental conditions that are not adequately captured by static terrain variables alone [44, 45].

Furthermore, the results highlight the

importance of interactions among conditioning factors. Landslide susceptibility is not controlled by a single variable but by the combined influence of topography, geology, hydrology, and land cover [46]. For example, steep slopes may remain stable under dense vegetation but become highly susceptible when vegetation is removed [45]. Similarly, weak geological formations exhibit higher failure potential when combined with high rainfall or proximity to faults [47, 48]. The ability of CatBoost to model such interactions is a major advantage over conventional approaches, which often assume independent or linear relationships among variables.

From an applied perspective, the susceptibility patterns provide valuable insights for hazard management and land-use planning. The identification of high-risk zones near infrastructure corridors and settlements underscores the role of human activities in modifying slope stability. Road construction, slope cutting, and deforestation significantly alter natural equilibrium conditions, increasing landslide susceptibility. The model outputs can therefore support decision-making by identifying areas where mitigation measures such as slope stabilization, drainage control, or vegetation restoration are required. In addition, the spatially explicit nature of the susceptibility map makes it suitable for integration into regional planning frameworks and disaster risk reduction strategies.

The results also demonstrate the broader applicability of machine learning approaches for landslide susceptibility assessment in data-scarce regions. The use of freely available Landsat 8 data, combined with GIS-based conditioning factors, provides a cost-effective and scalable solution for regional hazard mapping. This is particularly relevant for tropical mountainous regions where rapid environmental changes and limited field data pose challenges for traditional approaches.

However, several limitations remain that should be considered when interpreting the results. The model is based on static conditioning factors and does not explicitly incorporate temporal triggers such as rainfall intensity, duration, or antecedent moisture conditions. As a result, the susceptibility map represents potential rather than real-time hazard conditions. In addition, the quality and completeness of the landslide inventory influence model accuracy, and under-reporting of small or remote landslides may introduce uncertainty. The spatial resolution of input data, particularly Landsat imagery and DEM, may also limit the representation of fine-scale terrain features.

Overall, the findings confirm that integrating remote sensing data with advanced ensemble machine learning techniques significantly enhances landslide susceptibility modeling. The CatBoost–Landsat 8 framework provides a reliable and efficient approach for capturing complex environmental interactions and generating high-resolution susceptibility maps, offering valuable support for both scientific analysis and practical hazard management.

6. Conclusion

This study demonstrates the effectiveness of integrating Landsat 8–derived surface indicators with a CatBoost machine learning framework for landslide susceptibility mapping in a tropical monsoon mountainous region. The CatBoost model outperformed LR, SVM, MLP, and DNN in terms of AUC and predictive accuracy, confirming its strong capability to model complex nonlinear relationships among geo-environmental factors. The resulting susceptibility maps exhibit improved spatial reliability and enable more precise identification of high-risk zones.

The findings provide a practical basis for land-use planning, infrastructure development, and disaster risk reduction in Central Vietnam and similar terrains. However, model performance is dependent on the quality and completeness of landslide inventory data, and the use of medium-resolution Landsat imagery may limit fine-scale terrain representation. In addition, the current approach relies on static conditioning factors and does not explicitly incorporate temporal triggering mechanisms such as rainfall variability.

Future research should focus on integrating higher-resolution and multi-sensor datasets (e.g., LiDAR and SAR/InSAR), improving landslide inventory datasets, and incorporating dynamic variables such as rainfall thresholds and real-time monitoring. These improvements will support the development of more reliable and operational landslide early warning systems.

Acknowledgments

This research was financially supported by the Vietnam Academy of Science and Technology (VAST) under grant code number CBCLCA.16/25-27

References

[1] E.T.N. Kinyeru, M.N. Harris, J. Jansz. (2026). Why do landslides occur in Kenya, and what can be done to mitigate their occurrences?, *Natural Hazards*, 122, 155. https://doi.org/10.1007/s11069-025-07938-1

[2] A. Ramya, R. Poornima, G. Karthikeyan, S. Priyatharshini, K.G. Thanuja, P. Dhevagi. (2023). Climate-Induced and Geophysical Disasters and Risk Reduction Management in Mountains Regions. *Climate Change Adaptation, Risk Management and Sustainable Practices in the Himalaya, Springer, pp. 361-405.* https://doi.org/10.1007/978-3-031-24659-3_16

[3] T.A. Tuan, P.V. Hong, T. Tran Thi, N. Nguyen Thi Anh, D. Nguyen Van, H. Pham Tuan, P. Tran Van. (2024). Landslide susceptibility in Phuoc Son, Quang Nam: A deep learning approach. *Vietnam Journal of Earth Sciences*, 47(1), 39-57. https://doi.org/10.15625/2615-9783/21658

[4] T.V. Phong, H.B. Ly, P.T. Trinh, I. Prakash, D.T. Hoan. (2020). Landslide susceptibility mapping using Forest by Penalizing Attributes (FPA) algorithm based machine learning approach. *Vietnam Journal of Earth Sciences*, 42(3), 237-246. https://doi.org/10.15625/0866-7187/42/3/15047

[5] T.V. Phong, P.T. Truyen, N.V. Duong, N.L. Minh, N.D. Mau, I. Prakash, D.M. Duc, D.N. Duc. (2025). Landslide susceptibility mapping using Partial Decision Tree-Based hybrid artificial intelligence models. *Vietnam Journal of Earth Sciences*, 48(1), 145-166. https://doi.org/10.15625/2615-9783/24023

[6] V.C. Dat, P.T.T. Nga, T.H. Thai, T.V. Phong, M.V. Khiem, H.V. Dai, I. Prakash, B.T. Pham. (2025). Novel Integration of JCHAIDStar with ensemble techniques for comprehensive landslide and flash flood susceptibility mapping. *Vietnam Journal of Earth Sciences*, 48(2), 219-238. https://doi.org/10.15625/2615-9783/24030

[7] Y. Cui, L. Yang, C. Xu, J. Zheng. (2024). Spatial distribution of shallow landslides caused by Typhoon Lekima in 2019 in Zhejiang Province, China. *Journal of Mountain Science*, 21, 1564-1580. https://doi.org/10.1007/s11629-023-8377-y

[8] V. Pazzi, S. Morelli, R. Fanti. (2019). A Review of the Advantages and Limitations of Geophysical Investigations in Landslide Studies. *International Journal of Geophysics*, 2019, 2983087. https://doi.org/10.1155/2019/2983087

[9] D.D. Cham, N.V. Hong, N.N. Thang, N.T. Binh. (2023). The application of Google Earth Engine platform to assess the erosion - accretion processes of the estuary and coastal areas of Tra Khuc river, Quang Ngai province, Vietnam. *Vietnam Journal of Marine Science and Technology*, 23(3), 233-246. https://doi.org/10.15625/1859-3097/18634

[10] L.D. Hanh, H.T. Son, T.P. Tuan, V.H. Dang, N.N. Thang, B.A. Tuan, T.T. Binh, T.V. Nga, T.A. Tuan. (2025). Assessing the applicability of sentinel-2 remote sensing data to determine the bathymetry of a shallow coastal area: case study of Nhat Le Estuary, Quang Binh Province. *Vietnam Journal of Marine Science and*

Technology, 25(1), 17-28. https://doi.org/10.15625/1859-3097/22634

[11] B.T. Pham, V.D. Vu, R. Costache, T.V. Phong, T.Q. Ngo, T.-H. Tran, H.D. Nguyen, M. Amiri, M.T. Tan, P.T. Trinh, H.V. Le, I. Prakash. (2022). Landslide susceptibility mapping using state-of-the-art machine learning ensembles. *Geocarto International*, 37(18), 5175-5200. https://doi.org/10.1080/10106049.2021.1914746

[12] A. Roccati, G. Paliaga, F. Luino, F. Faccini, L. Turconi. (2021). GIS-Based Landslide Susceptibility Mapping for Land Use Planning and Risk Assessment. *Land*, 10(2), 162. https://doi.org/10.3390/land10020162

[13] R. Fell, J. Corominas, C. Bonnard, L. Cascini, E. Leroi, W.Z. Savage. (2008). Guidelines for landslide susceptibility, hazard and risk zoning for land use planning. *Engineering Geology*, 102(3-4), 85-98. https://doi.org/10.1016/j.enggeo.2008.03.022

[14] R.P. Quevedo, A. Velastegui-Montoya, N. Montalván-Burbano, F. Morante-Carballo, O. Korup, C. Daleles Rennó. (2023). Land use and land cover as a conditioning factor in landslide susceptibility: a literature review. *Landslides*, 20, 967-982. https://doi.org/10.1007/s10346-022-02020-4

[15] X. Yang, R. Liu, M. Yang, J. Chen, T. Liu, Y. Yang, W. Chen, Y. Wang. (2021). Incorporating Landslide Spatial Information and Correlated Features among Conditioning Factors for Landslide Susceptibility Mapping. *Remote Sensing*, 13(11), 2166. https://doi.org/10.3390/rs13112166

[16] L. Pathak, B. Baral, K. Joshi, D.R. Basnet, D. Godone. (2025). Landslides in the Himalayas: The Role of Conditioning Factors and Their Resolution in Susceptibility Mapping. *Geosciences*, 15(4), 131. https://doi.org/10.3390/geosciences15040131

[17] M. Shafapourtehrany, M. Batur, H. Özener, N. Lorestani, B. Kalantar, F. Shabani. (2026). Conventional and advanced geospatial techniques for landslide detection and modeling: a comprehensive overview. *Geoenvironmental Disasters*, 13, 3. https://doi.org/10.1186/s40677-025-00347-3

[18] P. Jiang, Z. Ma, G. Mei. (2026). Review article: Deep learning for potential landslide identification: data, models, applications, challenges, and opportunities. *Natural Hazards and Earth System Sciences*, 26(1), 487-529. https://doi.org/10.5194/nhess-26-487-2026

[19] C. Zhong, Y. Liu, P. Gao, W. Chen, H. Li, Y. Hou, T. Nuremanguli, H. Ma. (2020). Landslide mapping with remote sensing: challenges and opportunities. *International Journal of Remote Sensing*, 41(4), 1555-1581. https://doi.org/10.1080/01431161.2019.1672904

[20] N.S. Pawar, K.V. Sharma. (2025). Comprehensive review of remote sensing integration with deep learning in landslide forecasting and future directions. *Natural Hazards*, 121, 23687-23721. https://doi.org/10.1007/s11069-025-07171-w

[21] R. Sharf, Y. Sharf, K. Abbas, M. Alam. (2026). Environment, Geography, Landslide, and Sustainability: An Overview of Geography–Landslide Interface. *Advanced Landslide Susceptibility Modeling, Springer, pp. 3-24.* https://doi.org/10.1007/978-3-032-14485-0_1

[22] P.K. Badapalli, A.B. Nakkala, R.B. Kottala, S. Gugulothu, F.F.B. Hasher, V.N. Mishra, M. Zhran. (2025). Landslide Susceptibility Level Mapping in Kozhikode, Kerala, Using Machine Learning-Based Random Forest, Remote Sensing, and GIS Techniques. *Land*, 14(7), 1453. https://doi.org/10.3390/land14071453

[23] M. Scaioni, L. Longoni, V. Melillo, M. Papini. (2014). Remote Sensing for Landslide Investigations: An Overview of Recent Achievements and Perspectives. *Remote Sensing*, 6(10), 9600-9652. https://doi.org/10.3390/rs6109600

[24] V. Upadhyay. (2024). Landslide Hazard Risk and Vulnerability Monitoring—GIS Based Approach. *Landslide: Susceptibility, Risk Assessment and Sustainability. Application of Geostatistical and Geospatial Modeling, Advances in Natural and Technological Hazards Research, vol 52, Springer, pp. 53-86.* https://doi.org/10.1007/978-3-031-56591-5_3

[25] S. Mihu, K.K.S. Tomar, A. Kumar, P.P. Choudhari, A. Raju, M. Gentilucci, M. Barbieri, P. Kumar, R. Rongpi. (2026). Machine Learning-based Landslide Susceptibility Modeling in the Dibang Valley, NE India. *Earth Systems and Environment.* https://doi.org/10.1007/s41748-026-01036-3

[26] Q. Zhang, T. Wang. (2024). Deep Learning for Exploring Landslides with Remote Sensing and Geo-Environmental Data: Frameworks, Progress, Challenges, and Opportunities. *Remote Sensing*, 16(8), 1344. https://doi.org/10.3390/rs16081344

[27] D.A.M. Atok, S.S. Chai. (2025). Addressing overfitting and overestimation challenges in landslide susceptibility modeling: a case study of Penang Island, Malaysia. *Natural Hazards*, 121, 13577-13604. https://doi.org/10.1007/s11069-025-07329-6

[28] T. Kavzoglu, A. Teke. (2022). Predictive Performances of Ensemble Machine Learning Algorithms in Landslide Susceptibility Mapping Using Random Forest, Extreme Gradient Boosting (XGBoost) and Natural Gradient Boosting (NGBoost). *Arabian Journal for Science and Engineering*, 47, 7367-7385. https://doi.org/10.1007/s13369-022-06560-8

[29] E.K. Sahin. (2022). Comparative analysis of gradient boosting algorithms for landslide susceptibility mapping. *Geocarto International*, 37(9), 2441-2465. https://doi.org/10.1080/10106049.2020.1831623

[30] H. Jiang, Hao·Zhou, J. Wu, M. Liu, Y. Wu, Y. Guo. (2026). Ensemble learning for landslide susceptibility mapping: a review of machine learning and hybrid approaches. *Carbonates and Evaporites*, 41, 16. https://doi.org/10.1007/s13146-025-01221-x

[31] Y. Cai, Y. Yuan, A. Zhou. (2024). Predictive slope stability early warning model based on CatBoost. *Scientific Reports*, 14, 25727. https://doi.org/10.1038/s41598-024-77058-6

[32] R.S. Ajin, S. Segoni, R. Fanti. (2024). Optimization of SVR and CatBoost models using metaheuristic algorithms to assess landslide susceptibility. *Scientific Reports*, 14, 24851. https://doi.org/10.1038/s41598-024-72663-x

[33] J.T. Hancock, T.M. Khoshgoftaar. (2020). CatBoost for big data: an interdisciplinary review. *Journal of Big Data*, 7, 94. https://doi.org/10.1186/s40537-020-00369-8

[34] V.-H. Nhu, A. Mohammadi, H. Shahabi, B.B. Ahmad, N. Al-Ansari, A. Shirzadi, J.J. Clague, A. Jaafari, W. Chen, H. Nguyen. (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. https://doi.org/10.3390/ijerph17144933

[35] H. Ha, Q.D. Bui, D.T. Tran, D.Q. Nguyen, H.X. Bui, C. Luu. (2025). Improving the forecast performance of landslide susceptibility mapping by using ensemble gradient boosting algorithms. *Environment, Development and Sustainability*, 27, 18409-18443. https://doi.org/10.1007/s10668-024-04694-3

[36] Z. Zulfahmi, M.H.Z. Putra, D. Sarah, A. Tohari, N. Madiutomo, P. Hartanto, R. Damayanti. (2025). GIS-Based Landslide Susceptibility Mapping with a Blended Ensemble Model and Key Influencing Factors in Sentani, Papua, Indonesia. *Geosciences*, 15(10), 390. https://doi.org/10.3390/geosciences15100390

[37] A. Joshi, P. Saggar, R. Jain, M. Sharma, D. Gupta, A. Khanna. (2021). CatBoost — An

Ensemble Machine Learning Model for Prediction and Classification of Student Academic Performance. *Advances in Data Science and Adaptive Analysis*, 13(2), 2141002. DOI: 10.1142/S2424922X21410023

[38] D.V. Dao, A. Jaafari, M. Bayat, D. Mafi-Gholami, C. Qi, H. Moayedi, T.V. Phong, H.-B. Ly, T.-T. Le, P.T. Trinh, C. Luu, N.K. Quoc, B.N. Thanh, B.T. Pham. (2020). A spatially explicit deep learning neural network model for the prediction of landslide susceptibility. *CATENA*, 188, 104451. https://doi.org/10.1016/j.catena.2019.104451

[39] B.T. Pham, D.T. Bui, I. Prakash, M.B. Dholakia. (2017). Hybrid integration of Multilayer Perceptron Neural Networks and machine learning ensembles for landslide susceptibility assessment at Himalayan area (India) using GIS. *CATENA*, 149 (Part 1), 52-63. https://doi.org/10.1016/j.catena.2016.09.007

[40] S. Suthaharan. (2016). Support Vector Machine. *Machine Learning Models and Algorithms for Big Data Classification: Thinking with Examples for Effective Learning. Integrated Series in Information Systems, vol 36. Springer, pp. 207-235.* https://doi.org/10.1007/978-1-4899-7641-3_9

[41] F. He, F. Cai. (2026). Feature Selection and Comparison of Logistic Regression and Random Forest for Stability Assessment of Landslide Dams. *Landslides*, 23, 1149-1161. https://doi.org/10.1007/s10346-025-02677-7

[42] M. Han, Y. Yang. (2026). Research on interpretable machine learning modeling and spatial prediction of cross-landform landslide disaster mechanism: A case study of Shaanxi Province. *PLOS ONE*, 21(5), e0347996. https://doi.org/10.1371/journal.pone.0347996

[43] M. Safabakhshpachehkenari, H. Tonooka. (2023). Assessing and Enhancing Predictive Efficacy of Machine Learning Models in Urban Land Dynamics: A Comparative Study Using Multi-Resolution Satellite Data. *Remote Sensing*, 15(18), 4495. https://doi.org/10.3390/rs15184495

[44] K.C. Niraj, A. Singh, D.P. Shukla. (2023). Effect of the Normalized Difference Vegetation Index (NDVI) on GIS-Enabled Bivariate and Multivariate Statistical Models for Landslide Susceptibility Mapping. *Journal of the Indian Society of Remote Sensing*, 51, 1739-1756. https://doi.org/10.1007/s12524-023-01738-5

[45] E.B. Masi, S. Segoni, V. Tofani. (2021). Root Reinforcement in Slope Stability Models: A Review. *Geosciences*, 11(5), 212. https://doi.org/10.3390/geosciences11050212

[46] M.J. Crozier, T. Glade. (2005). Landslide Hazard and Risk: Issues, Concepts and Approach. *Landslide Hazard and Risk*, pp. 1-40.https://doi.org/10.1002/9780470012659.ch1

[47] M.J. Crozier. (1985). Landslides: causes, consequences and environment. *Routledge*.

[48] F.C. Dai, C.F. Lee, Y.Y. Ngai. (2002). Landslide risk assessment and management: an overview. *Engineering Geology*, 64(1), 65-87. https://doi.org/10.1016/S0013-7952(01)00093-X