Environment and Natural Resources Journal

Volume 24
Issue 2
2026

Forest Cover and Landslide Susceptibility Assessment Using a Machine Learning Approach in Northern Midland and Mountainous Region of Vietnam

Thuong Tran[1], Hoa Trieu[2*], and Nathaniel Bantayan[3]

[1]*International School - Thai Nguyen University, Vietnam*
[2]*Thai Nguyen University of Agriculture and Forestry, Vietnam*
[3]*University of the Philippines Los Baños, Philippines*

ARTICLE INFO

Received: 16 Jun 2025
Received in revised: 11 Nov 2025
Accepted: 13 Nov 2025
Published online: 15 Jan 2026
DOI: 10.32526/ennrj/24/20250178

Keywords:
Forest/ Landslides susceptibility/ ML/ KNN/ RF/ MLP/ Cau River Watershed

*** Corresponding author:**
E-mail:
xuanhoatrieu@tuaf.edu.vn

ABSTRACT

Landslides are a major geo-environmental hazard in Vietnam's midland and mountainous regions, further intensified by land-use pressures and climate change. This study investigated the influence of forest cover on landslide susceptibility in Cau River Watershed. A forest status map was constructed using inventory and field data by the K-Nearest Neighbors (KNN) algorithm, while landslide susceptibility was modeled using historical events and nine conditioning factors through a hybrid machine learning approach integrating Random Forest (RF), Multilayer Perceptron (MLP) and KNN. The proposed hybrid model achieved an overall accuracy of 85.33%, demonstrating its robustness in susceptibility prediction. Results indicated that natural and native-species forests significantly reduce landslide density and susceptibility relative to non-forested areas and exotic plantations. These findings highlight the critical role of forest structure and species composition in stabilizing slopes. The study provides evidence-based insights to guide adaptive land management, forest policy, and regional strategies for climate resilience and sustainable development.

HIGHLIGHTS

A hybrid KNN-RF-MLP framework was developed for landslide susceptibility mapping. The hybrid model achieved 85.33% accuracy, outperforming the standard RF model. Forest cover maps were enhanced using KNN imputation for incomplete field data. Native forests and species-rich stands greatly reduced landslide susceptibility. Findings guide sustainable reforestation and climate-resilient land management.

1. INTRODUCTION

Understanding the link between landslide susceptibility and climate-induced hazards is increasingly critical, particularly in Southeast Asia, where climate variability is intensifying. Northern Vietnam is characterized by steep terrain, high rainfall, and rapid land-use changes. In such regions, assessing landslide susceptibility is vital for climate resilience. It enables the identification of high-risk zones and supports sustainable land-use planning (Chen and Pan, 2019).

Forests and vegetation cover are essential in mitigating the impacts of natural hazards, particularly landslides (WB, 2019). Forest ecosystems provide essential services such as slope stabilization, erosion control, water regulation, and biodiversity conservation. Tree roots reinforce soil structure and improve infiltration, while canopy cover reduces the erosive impact of rainfall (FAO, 2013). These functions make intact and biodiverse forests key to reducing landslide severity and protecting human settlements and infrastructure.

In Vietnam, recent government initiatives have emphasized reforestation and early zoning of landslide-prone areas as part of national adaptation strategies to natural disasters (VNDMA, 2023b). The importance of forest-based resilience is also supported by regional studies. According to FAO's report (2013), Asia remains one of the most landslide-prone continents, where increasing rainfall extremes and

Citation: Tran T, Trieu H, Bantayan N. Forest cover and landslide susceptibility assessment using a machine learning approach in northern midland and mountainous region of Vietnam. Environ. Nat. Resour. J. 2026;24(2):209-221. (https://doi.org/10.32526/ennrj/24/20250178)

ongoing land degradation continue to exacerbate disaster risks. In mountainous provinces of northern Vietnam, catastrophic landslides in recent years have resulted in significant loss of life and damage to both agricultural and forested land (VNDMA, 2023b).

To improve hazard prediction and environmental planning, machine learning (ML) techniques are increasingly used in landslide susceptibility modeling. These data-driven approaches can integrate diverse environmental variables to generate highly accurate susceptibility maps, even in regions with limited or incomplete field data. Such maps are essential tools for identifying zones exposed to both environmental and meteorological risks. They provide spatially explicit information that facilitates early warning systems, risk communication, and adaptive land-use planning.

In northern Vietnam, heavy rainfall is widely recognized as the dominant triggering factor of landslides (Le and Kaneko, 2017). However, forest cover is a critical variable that can be directly managed to mitigate slope instability. With its relatively high forest cover, Cau River Watershed offers an ideal setting to examine how forest classifications influence landslide susceptibility. Therefore, this study aims to: (i) assess the spatial distribution of forest cover and landslide susceptibility in Cau River Watershed; (ii) develop and validate a hybrid machine learning framework (KNN-RF-MLP) for improving landslide susceptibility prediction; and (iii) evaluate the effects of different forest types on landslide occurrence.

2. METHODOLOGY

2.1 Study area

This study focuses on Cau River Watershed located in Thai Nguyen Province (CRWTN), as illustrated in Figure 1. The watershed spans from latitude 21°26′8″N to 22°2′54″N and longitude 105°28′36″E to 106°7′41″E, covering approximately (3,527 km², equivalent to 79.6%) of the total land area of Thai Nguyen Province (as of June 2025). Of this area, around 1,876 km² is classified as forest land (VNGSO, 2024).

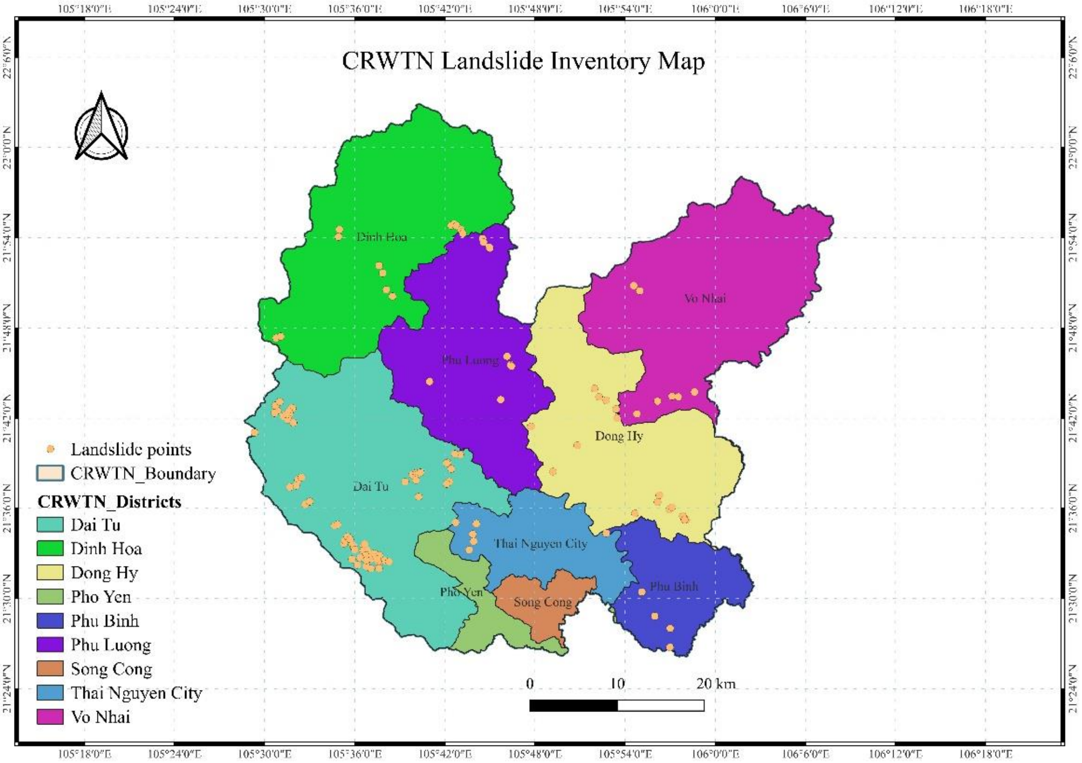


Figure 1. Landslide inventory map of CRWTN

The study area is characterized by typical natural features of a mountainous province in northern Vietnam, situated within a tropical monsoon climate zone. The topography is predominantly composed of hills and mountains, which define the region's dominant terrain. Most of the area lies below 300 meters in elevation, although approximately two-thirds of the province's total land area consists of hilly and mountainous terrain exceeding 100 meters in elevation (Le and Kaneko, 2017). The region experiences a tropical monsoon climate with a distinctly cold winter. Annual precipitation is high, with a prolonged rainy season extending from May to October. The most intense rainfall typically occurs between June and August, during which maximum daily rainfall events range from approximately 208 mm to 496 mm. This contributes to an annual rainfall total ranging from 1,360 to 2,572 mm. According to annual forest reports by the Vietnam Administration of Forestry (VAF, 2024), the forested area in Thai Nguyen Province has shown a general increasing trend from 2010 to 2023. As of 2023, forest cover in Cau River Watershed exceeded 1,575 km^2, representing 56.1% of the total study area (TNFPD, 2024). Situated at the heart of the Northern Midland and Mountainous Region - one of Vietnam's most landslide-prone areas (Bui, 2017), (Le and Kaneko, 2017) - the study area has experienced significant landslide events. Over the past two decades, Thai Nguyen Province has suffered extensive damage from landslides, resulting in 110 fatalities and widespread destruction of residential structures (VNDMA, 2023a).

2.2 Data and methods

The methodological framework of this study consisted of: (i) data collection and preprocessing, (ii) forest status and landslide susceptibility mapping, and (iii) empirical analysis of the relationships between forest and landslides (Figure 2).

2.2.1 Data collection and processing

The mapping component involved the delineation of forest types and the spatial distribution of areas susceptible to landslides. For forest status mapping, data were collected from the Thai Nguyen Forest Protection Department, including official forest inventory records. Additional field data were obtained from the sample plots.

For landslide susceptibility modeling, the input dataset included a landslide inventory map (Figure 1) that was developed using historical landslide events from previous research projects, field observations, and entries from the NASA Global Landslide Catalog (updated in 2019), as well as various landslide conditioning factors (LCFs) as in Figure 3. These nine LCFs were selected to represent the topographic, geological, vegetation, land cover, and climatic characteristics of the study area (Table 1).

Table 1. Landslide conditioning factors

Factor(s)	Category	Source/Resolution	Period
Slope, Elevation, Drainage Density, Relief Degree of Land Surface (RDLS)	Topographic	DEM (SRTM 30 m, USGS)	2020
Lithology	Geological	Thai Nguyen Department of Science and Technology	2015
Soil type	Pedological	National Pedology Map (Open Development Mekong)	2015
NDVI	Vegetation	Landsat 8 (30 m, USGS)	2023
LULC	Land cover	JAXA Northern Vietnam LULC Map	2020
Precipitation	Climatic	Hydrometeorological stations, Thai Nguyen	2010-2023

These factors have been widely recognized in previous studies as key variables influencing slope stability and landslide occurrence.

2.2.2 Forest status and landslide susceptibility mapping

a) Forest status mapping

To establish a forest status map, data from the Forest Inventory provided by the local Forest Protection Department were integrated with field survey records. A significant challenge encountered during this process was the presence of numerous locations with incomplete or missing data. To address this limitation, the K-Nearest Neighbors (KNN) algorithm was applied to validate existing entries and impute the missing values, thereby enhancing the completeness and reliability of the dataset.

KNN algorithm is a widely used, straightforward, and effective supervised machine learning technique. It classifies new data points based on the majority class among the K closest instances in the training dataset, using a distance metric such as Euclidean distance (Harrison, 2018).

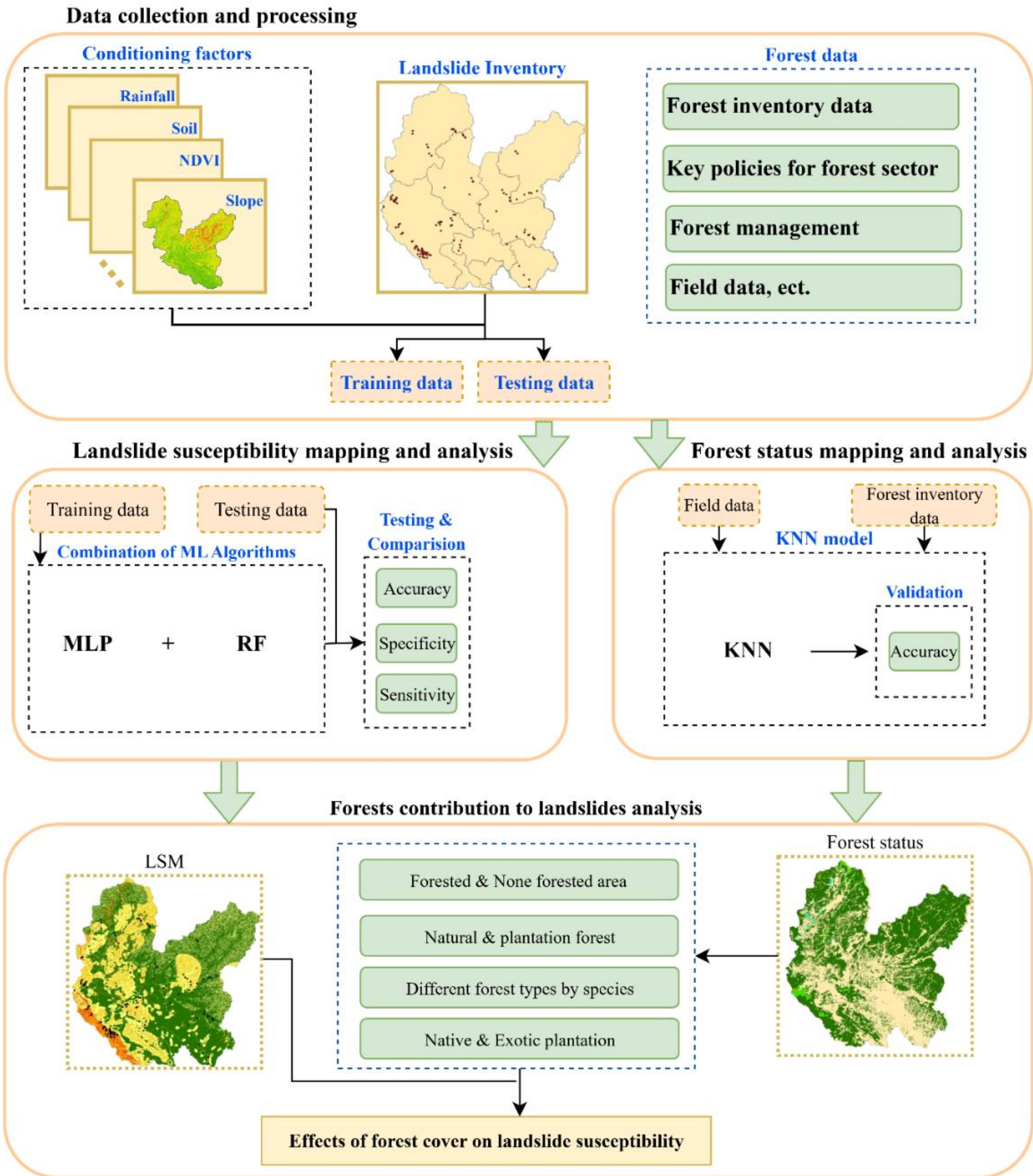


Figure 2. Framework for Forest Cover and Landslide Susceptibility Assessment

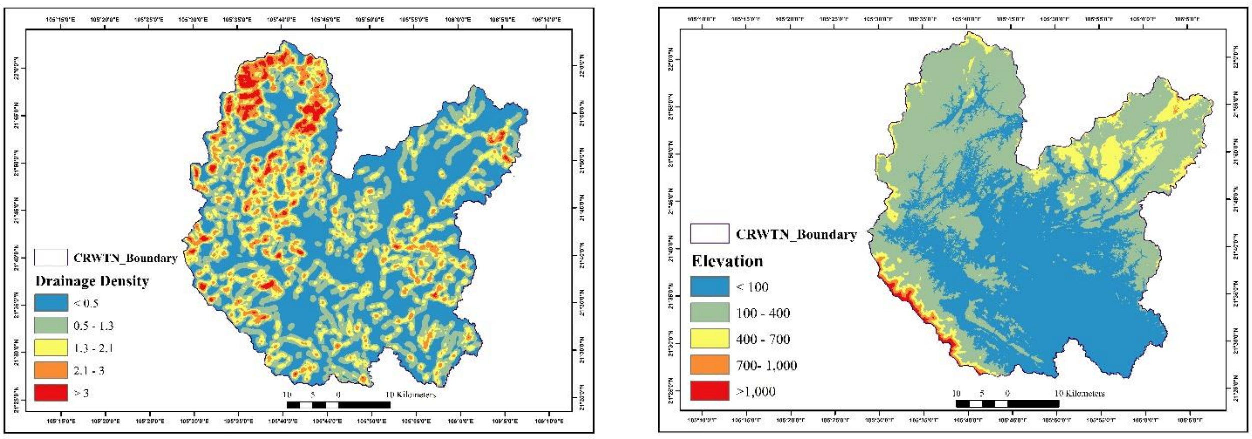


Figure 3. Landslide conditioning factors (LCFs)

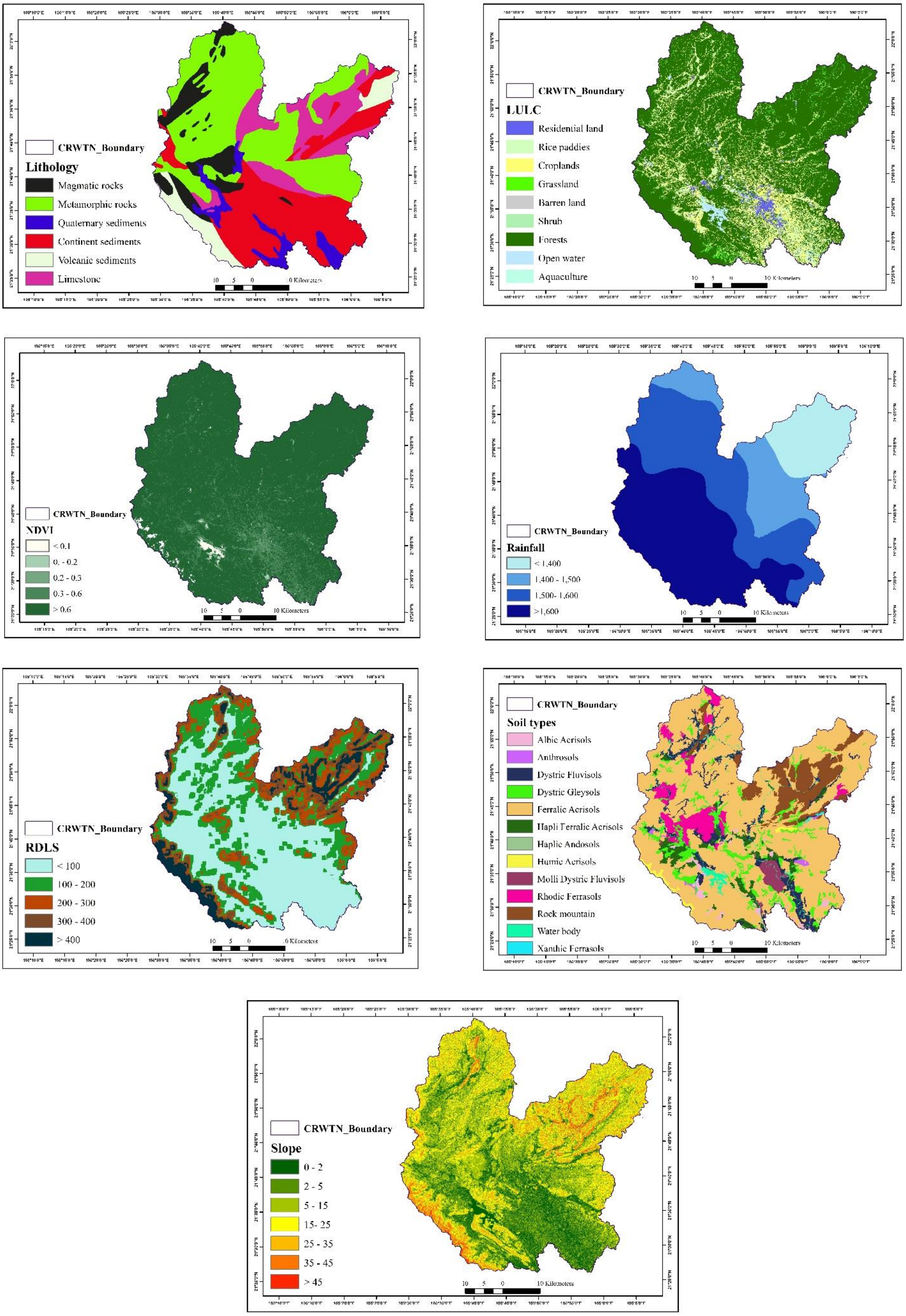
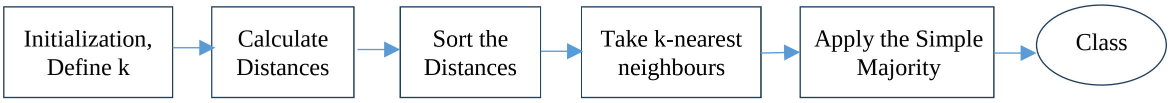


Figure 3. Landslide conditioning factors (LCFs) (cont.)

Basic steps of how the KNN works are depicted as below:



The performance of the KNN model has been evaluated using the accuracy criteria, as it provides a straightforward and comprehensive measure of overall classification correctness. Given that the dataset was relatively balanced and KNN was applied primarily for imputation rather than predictive modeling, accuracy (ACC) was considered sufficient for assessing model quality. This quality criterion refers to the closeness of a measurement to the true or accepted value (James, 2013).

$$\mathrm{ACC} = \frac{\mathrm{TP+TN}}{\mathrm{TP+FP+TN+FN}}$$

The model's output has been imported into a GIS tool. Subsequently, a forest status map for the study area was generated by integrating these processed data.

b) Landslide susceptibility mapping

Step 1: Landslide susceptibility model construction

For landslide susceptibility modeling, the input dataset contained a landslide inventory map documenting 125 landslide events and nine LCFs (Figure 3). All grid cells in the dataset referring to landslides were assigned a value of 1. An equal number of grid cells referring to non-landslide was randomly sampled from the landslide-free regions and assigned a value of 0. Corresponding LCF values were extracted to form labeled datasets indicating the presence or absence of landslides, including all LCF attributes. The dataset was randomly divided into training and testing subsets using a 70/30 split. To ensure robust classification performance, a fivefold cross-validation procedure was employed. The dataset was partitioned into five subsets, with the model trained on four and tested on the remaining one. This process was repeated five times to provide a reliable estimate of model performance.

A two-phase training strategy (Tran et al., 2024) was used for model construction. In Phase 1, a ML model was trained on the original training dataset. The resulting outputs were then integrated with the original data to create an augmented dataset. In Phase 2, the model was retrained using this enhanced input, and its predictive performance was assessed. This strategy effectively leveraged the Phase 1 model to enrich the training data, while the Phase 2 model performed the final landslide susceptibility prediction. In this study, a hybrid approach combining Random Forest and Multilayer Perceptron was developed and applied for landslide susceptibility mapping. Additionally, the KNN algorithm was applied to verify and impute missing data before executing the two main phases (Figure 4).

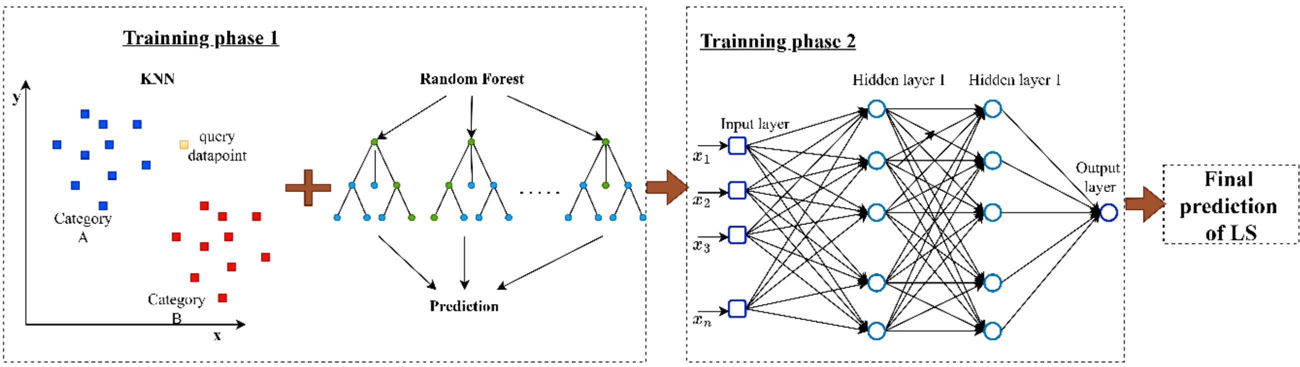


Figure 4. Architecture of the hybrid ML model for landslide susceptibility prediction

- *Random forest algorithm*

RF algorithm employs an ensemble of decision trees to enhance the accuracy and robustness of predictions (Breiman, 2001). Each tree in the ensemble is trained on a randomly selected subset of the training data and a random subset of input features, which helps reduce the inter-tree correlation and improves the model's generalization capability. In this work, to further optimize the performance of the RF model, hyperparameter tuning was performed using the grid search technique. Hyperparameters are defined as user-set parameters that influence the learning process but are not learned directly from the data. Grid search systematically explored a predefined range of hyperparameter values to identify the combination that yielded the best performance on a validation set. Key hyperparameters that were tuned in RF models include: (i) the number of decision trees in the ensemble; (ii) the maximum depth of each tree; and (iii) the maximum number of features considered when splitting a node.

- *Multilayer perceptron*

MLP is a type of feedforward artificial neural network that consists of three main components: an input layer, one or more hidden layers, and an output layer. The input layer receives variables representing the features of the dataset. Each neuron (or node) in a hidden or output layer computes a weighted sum of the inputs it receives from the previous layer, adds a bias term, and applies a nonlinear activation function, such as the rectified linear unit used in this study, to introduce non-linearity into the model.

During training, MLP learns by adjusting the weights and biases of the network through the backpropagation algorithm, in conjunction with an optimization technique such as stochastic gradient descent in this work. The training objective is to minimize a loss function - commonly mean squared error for regression or cross-entropy for classification - by computing gradients of the loss with respect to each weight and bias and updating them iteratively. The architecture of MLP network used in this study is illustrated in Figure 4 (phase 2).

Step 2: Model validation

Several statistical metrics-namely Accuracy (ACC), Sensitivity (SEN), and Specificity (SPE) - were employed to evaluate the performance of the predictive models. Accuracy represents the overall correctness of the model by measuring the percentage of total predictions (both positive and negative) that are correct. It indicates how closely the model's predictions match with the actual outcomes. Sensitivity, also known as the true positive rate or recall, measures the model's ability to correctly identify positive instances (e.g., actual landslides). Specificity evaluates the model's ability to correctly classify negative instances (e.g., non-landslide areas). These evaluation metrics were calculated using the following formulas:

$$ACC = \frac{TP+TN}{TP+FP+TN+FN}; \quad SEN = \frac{TP}{TP+FN}; \quad SPE = \frac{TN}{TN+FP}$$

Where: TP (True Positive): the percentage of actual landslide cases correctly classified as landslides; FP (False Positive): the percentage of non-landslide cases incorrectly classified as landslides; TN (True Negative): the percentage of non-landslide cases correctly classified as non-landslides; FN (False Negative): the percentage of actual landslide cases incorrectly classified as non-landslides.

Step 3: Map establishment: The output of the model was imported into a GIS tool to create a landslide susceptibility map.

2.2.3 Effect of forest cover on landslides assessment

Using the forest status map and the landslide susceptibility map generated in the previous sections, the relationship between forest cover and landslide occurrence was evaluated. The analysis considered the distribution and susceptibility of landslides across the following categories: (i) forested versus non-forested areas; (ii) broad forest classifications, including natural forests and planted forests (based on original classification criteria); (iii) forest types distinguished by species composition, such as woody forests, mixed-species forests, palm and coconut plantations, and bamboo forests; (iv) subcategories of planted forests, specifically native versus exotic plantations.

Tools: Python's scikit-learn, Google Earth Engine, ArcGIS 10.5.

3. RESULTS

3.1 Forest status mapping and analysis in CRWTN

A KNN model was developed to generate the forest status map of CRWTN, utilizing data from the Local Forest Protection Department's forest inventory and supplementary field observations. Missing values in the dataset were imputed using predictions generated by the trained model. The model's performance was tested over an area of 2,423.85 km² in Van Yen commune, accounting for 32.37% of the total area with initially missing data, and achieved an accuracy of 86.26%. As a result, in Van Yen commune, forested land covers 86.41% of the total area, with a total forest area of 2,097 hectares. This includes 1,059.66 hectares of natural forests and 1,037.34 hectares of planted forests. Based on species classification, the commune contains 1,833.66

hectares of woody forests, 5.13 hectares of bamboo forests, and 258.21 hectares of mixed forests. After validation, the model was applied to the entire study area to produce the final forest status map of CRWTN (Figure 5).

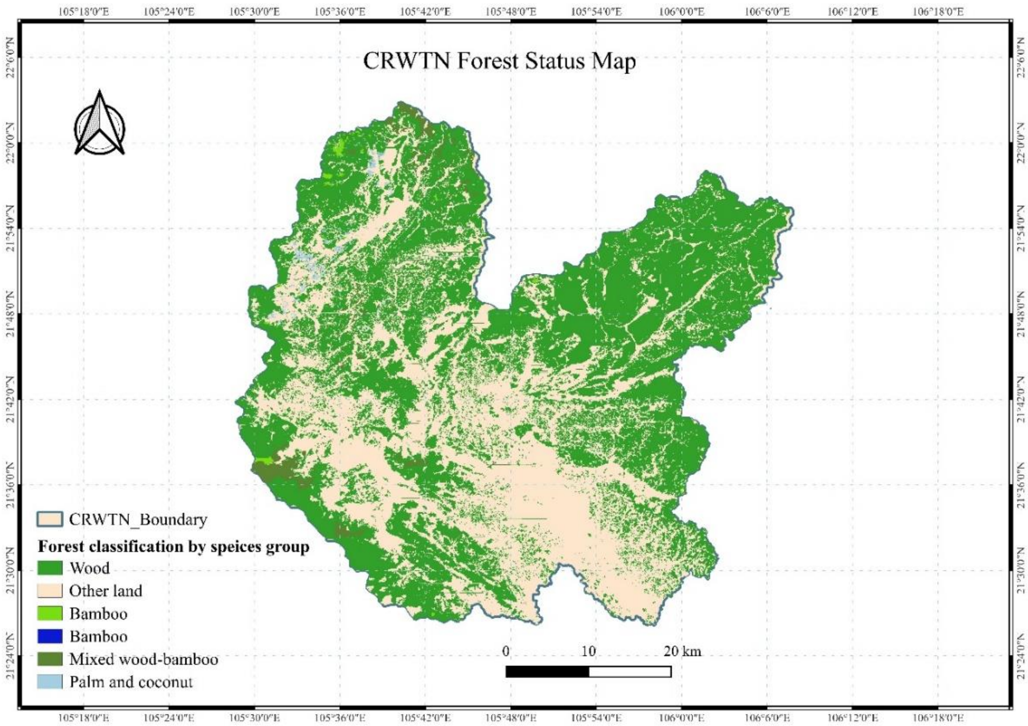


Figure 5. Forest status map in terms of species

Regarding forest origin, among the total 158,515.59 hectares of forested land in the study area, natural forests account for 40.8%, while planted forests represent 59.2%. Based on species classification, four main forest types are identified: woody forests (151,708.7 ha), bamboo forests (807.66 ha), mixed forests (4,467.06 ha), and palm and coconut plantations (1,234.19 ha). Woody forests are the most dominant, covering 95.89% of the total forest area in the watershed. In addition to native tree species, exotic species are also widely used in planted forests throughout CRWTN, primarily for the timber production and other economic purposes.

Similar to other forest-dominated regions in Vietnam, forests in CRWTN play a vital role in the local environment, society, and economy. They provide essential ecosystem services such as water supply, soil protection, flood control, air purification, and biodiversity conservation. Forests also support key industries including timber, agriculture, tourism, and food processing. Valuable species like pine, oak, and rosewood are used for wood products, while fruit trees and medicinal plants support local livelihoods. Therefore, effective forest protection and management are crucial for the sustainable development of CRWTN and the well-being of its communities.

3.2 Landslide susceptibility mapping and analysis in CRWTN

The two-phase training approach was used to develop a hybrid model integrating KNN, RF, and MLP. To evaluate the effectiveness of the proposed hybrid model, its performance was compared with the conventional RF model using three common classification metrics: ACC, SEN, and SPE. These metrics provide a comprehensive assessment of the model's ability to correctly classify landslide-prone and non-prone areas. As shown in Table 2, the hybrid model outperformed the single RF model across all evaluation criteria.

Table 2. The performance of models

Model	ACC	SEN	SPE
RF	76%	76%	76%
Hybrid	85.33%	85.71%	83.33%

Based on its superior classification performance, the hybrid model was applied to generate the landslide susceptibility map for CRWTN. The resulting map provides a spatial representation of landslide risk levels across the region and was divided into five susceptibility categories (Bui, 2017), (Chen and Pan, 2019): very low, low, moderate, high, and very high (Figure 6). This classification scheme enabled more effective interpretation of risk zones and supports the development of targeted mitigation and land-use planning strategies.

The spatial distribution of landslide susceptibility in CRWTN showed a predominance of low-risk zones. Areas classified as very low susceptibility covered approximately 154,129 hectares, accounting for 54.88% of the total area. Low susceptibility zones represented 48,334 hectares (17.21%), while moderate susceptibility areas spanned 22,159 hectares (7.89%). In contrast, regions of high and very high susceptibility were more limited in extent, comprising 14,258 hectares (5.08%) and 41,973 hectares (14.94%), respectively. These results indicate that approximately 20% of the area was classified as having high to very high susceptibility, emphasizing the need for detailed monitoring and proactive risk management in those zones.

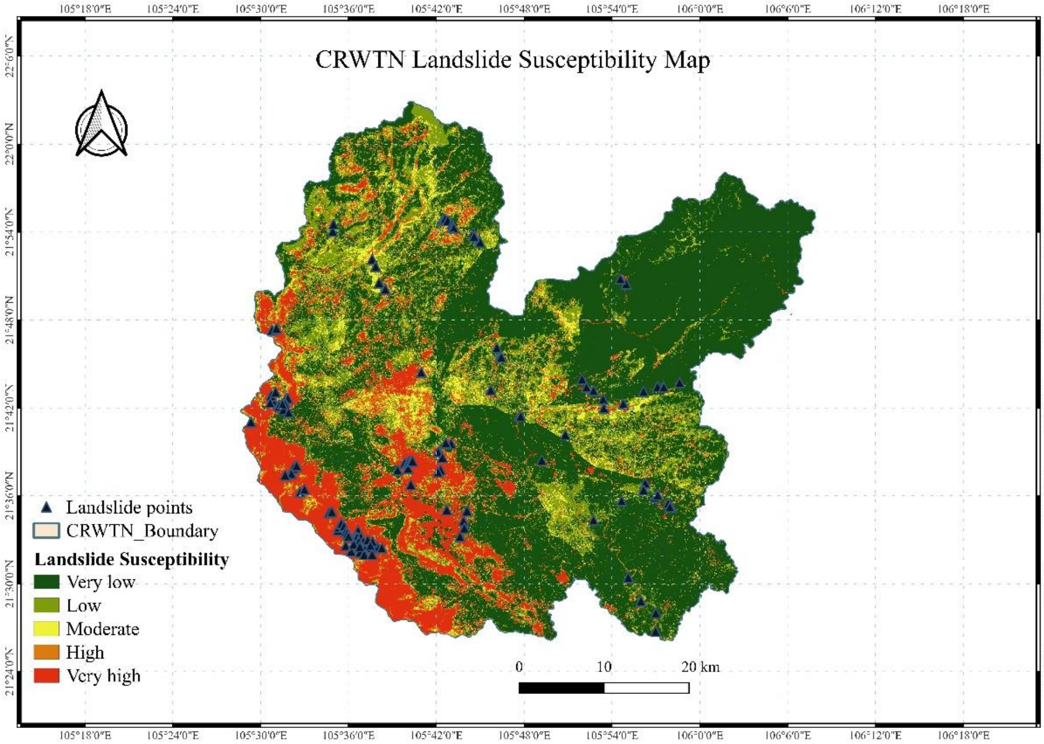


Figure 6. Landslide susceptibility map produced by the hybrid model

3.3 Effects of forest cover on landslide occurrences in CRWTN

The multi-dimensional assessment employed in this study enabled a deeper understanding of whether forest cover contributes to reducing landslide occurrence or, conversely, is associated with heightened landslide susceptibility in the study area. Using the forest status map and the landslide susceptibility map generated in previous sections, the relationship between forest cover and landslide occurrence was first assessed through the calculation of landslide density across various land cover categories. This assessment considered both forested and non-forested areas, as well as different forest classifications. These classifications included: (i) origin-based categories, such as natural forests and planted forests; (ii) species-based types, including woody forests, mixed forests, palm and coconut plantations, and bamboo forests; and (iii) planted forest types, distinguishing between native and exotic plantations (Figure 7). In this context, landslide density is defined as the number of landslides occurring per square kilometer within a specific land cover type. This metric was employed to evaluate the spatial variation in landslide occurrence across the identified forest categories and to provide insights into the potential mitigating effects of different forest types on landslide susceptibility.

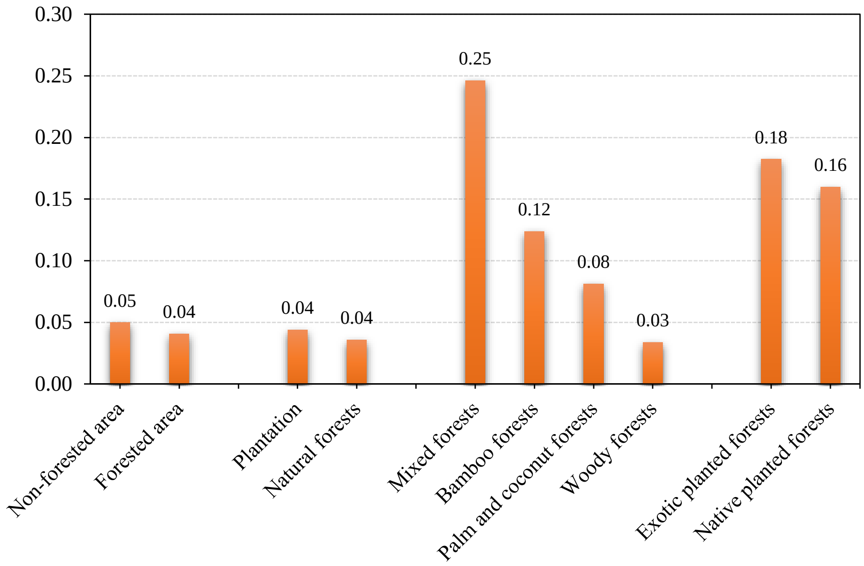


Figure 7. Landslide density across forest types

Furthermore, landslide susceptibility was analyzed and compared across multiple forest classifications, including forested versus non-forested areas, natural versus planted forests, four forest species groups, and native versus exotic plantations (Figures 8 and 9). In this context, landslide susceptibility refers to the quantitative or qualitative assessment of the likelihood and spatial distribution of either existing or potential landslide occurrences.

The analysis revealed that landslide density and susceptibility were both higher in non-forested areas than forested ones. As shown in Figure 8, non-forested land had 0.37% more area classified as high susceptibility and 3.65% more area classified as very high susceptibility compared to forested land. This demonstrates that forests play an important role in mitigating landslide risk, particularly by reducing the likelihood of landslides in high- and very high-susceptibility zones. These findings are consistent with previous studies highlighting the protective function of forests in slope stabilization (Schmaltz et al., 2017), (Chen and Shen, 2023), (Murgia et al., 2024), etc. Forested areas, through tree root systems and vegetation cover, contribute to soil stabilization and reduce surface runoff, whereas non-forested areas - especially those altered by human activities like deforestation or construction - are more vulnerable due to exposed and destabilized soils.

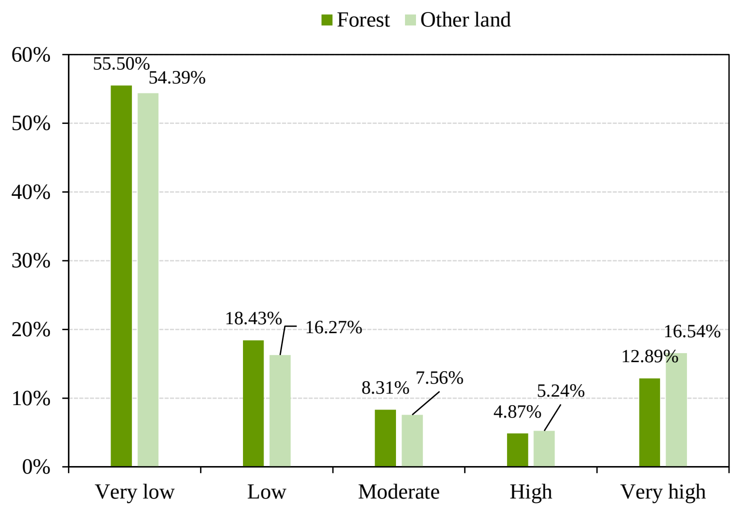


Figure 8. Landslide susceptibility on forest and non-forest areas

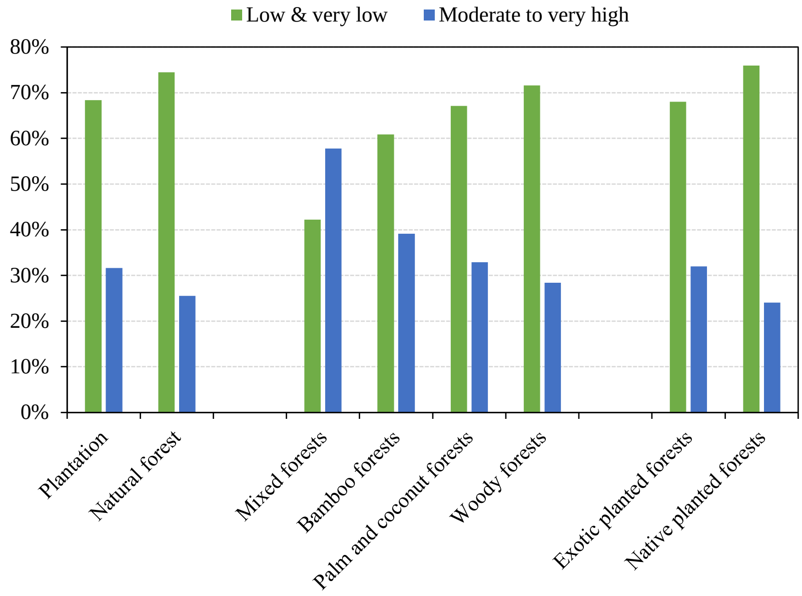


Figure 9. Landslide susceptibility over different forest types

Among forest types, natural forests exhibited lower landslide density and susceptibility than planted forests, likely due to their complex root structures, greater biodiversity, and higher organic matter content. In contrast, planted forests- often monocultures - tend to have simpler root systems and less stable soils. However, the impact of planted forests varied: native plantations demonstrated lower landslide density and higher proportions of very low to low susceptibility, suggesting better adaptation to local conditions and stronger soil-binding roots. Exotic plantations, on the other hand, were more prone to landslides, potentially due to shallow root systems and their influence on soil properties.

In the study area, native plantations are primarily composed of species such as *Manglietia conifera*, *Michelia mediocris*, and *Chukrasia tabularis*. These species are well adapted to local soil and climatic conditions and develop deep, fibrous root systems that contribute to slope stability. By contrast, exotic plantations are dominated by fast-growing species such as *Acacia mangium*, *Acacia auriculiformis*, and *Eucalyptus camaldulensis*, which are valued for timber and pulp production but typically form shallow root systems, providing limited reinforcement against slope failure (Bui, 2017; FAO, 2013).

When classified by species groups, woody forests had the lowest landslide occurrence and susceptibility, while mixed forests, bamboo forests, and palm and coconut plantations showed higher values. The shallow root systems common in these types likely provide less effective soil stabilization.

These findings confirmed the critical role of forest cover in reducing landslide risk in CRWTN. The effectiveness of forests depends on both forest types and species composition. In particular, well-managed forests - especially natural and native-species forests - can significantly contribute to landslide mitigation.

4. DISCUSSION

The integration of machine learning techniques in this study has proven valuable for analyzing the complex interactions between environmental variables and landslide susceptibility. The hybrid model addressed data gaps, improved classification accuracy, and enabled spatially explicit predictions of landslide-prone areas. This approach offers a replicable methodology for other regions. Given appropriate landslide inventory and conditioning data, it can be adapted and applied to other landslide-prone regions with similar environmental and data conditions, thereby extending its utility beyond CRWTN.

Previous studies have demonstrated the effectiveness of hybrid machine learning models in landslide susceptibility assessment, such as the Naïve Bayes-Random Subspace hybrid (Pham et al., 2021), the RF-MLP ensemble with Rotation Forest (Bui et al., 2022), and recurrent neural network frameworks (Wang et al., 2022). Beyond hybrid ML techniques, other studies examined the role of forests in slope stability: the spatio-temporal analysis showed that forest cover reduces shallow landslides in Switzerland

(Schmaltz et al., 2017); RF proved effective in forest-covered areas of Lin'an, China (Chen and Shen, 2023); and land-cover change simulations demonstrated that forest loss increases shallow landslide susceptibility in Central Italy (Murgia et al., 2024). Collectively, these studies reinforce the validity of our proposed KNN-RF-MLP framework and highlight the novelty of linking hybrid ML modeling with forest cover types to assess landslide susceptibility.

Beyond the core findings on forest effects, this discussion has extended to broader implications for land-use planning and disaster risk management. The study underscores the importance of not only preserving forest cover but also considering forest composition and management practices. While forest extent is important, the type of vegetation and its ecological characteristics play a critical role in influencing slope stability. These insights are relevant to Vietnam's reforestation programs. Such programs often favor fast-growing exotic species for economic purposes, but this can compromise long-term slope stability and resilience.

Nevertheless, the study had some limitations. Anthropogenic activities such as road construction, quarrying, and unplanned settlements substantially increase slope instability by altering drainage patterns, removing vegetation, and disturbing soil and rock structures. These factors can amplify landslide risk and, in some cases, override the stabilizing effects of forest cover—for example, slope undercutting along roads or excavation for quarrying often triggers failures even in forested areas. Because such activities were not explicitly included in the model, the results may underestimate their impact. Moreover, temporal changes in forest cover and landslide dynamics were not captured, emphasizing the need for long-term monitoring. Incorporating high-resolution temporal data, hydrological modeling, and socio-economic drivers in future research would provide a more comprehensive understanding of landslide processes and forest-landscape interactions.

Despite these limitations, the application of the hybrid machine learning framework has proven effective in capturing the multifaceted relationships between biophysical variables and landslide risk. The combined use of KNN for data imputation and RF-MLP for susceptibility modeling delivers robust performance and improves predictive accuracy.

5. CONCLUSION

This study developed a hybrid machine learning framework to assess landslide susceptibility in Cau River Watershed by integrating forest status mapping with spatial analysis of environmental conditions. The findings have demonstrated the added value of combining KNN, RF, and MLP algorithms for imputing missing data and improving the accuracy of susceptibility prediction. By linking forest classifications with landslide occurrence, the research provides practical insights to support evidence-based forest and land-use management. The results advocate for the prioritization of natural forests and native-species plantations in reforestation and conservation strategies, given their stronger slope-stabilizing functions. For policymakers, planners, and environmental managers, this work reinforces the importance of integrating ecological considerations into spatial planning and disaster risk reduction frameworks.

AUTHOR CONTRIBUTIONS

Data Collection and Experimental Run, Thuong Tran and Hoa Trieu; Methodology, Validation, Supervision and Original Draft Writing, Thuong Tran, Hoa Trieu and Nathaniel Bantayan; Formal Analysis, Thuong Tran; Data Curation, Visualization, and Reviewing and Editing of the manuscript, Thuong Tran and Hoa Trieu and Nathaniel Bantayan.

DECLARATION OF CONFLICT OF INTEREST

The authors declare that they have no conflicts of interest.

REFERENCES

Breiman L. Random forests. Machine Learning 2001;45:5-32.

Bui DT. Spatial prediction of rainfall-induced landslides for the Lao Cai Area (Vietnam) using a hybrid intelligent approach of least squares support vector machines inference model and artificial bee colony optimization. Landslides 2017;14:447-58.

Bui DT, Shirzadi A, Shahabi H, Pham BT, Singh SK, Chapi K, et al. A novel integrated approach of random forest-multilayer perceptron with rotation forest for landslide susceptibility assessment. Environmental Earth Sciences 2022;81:Article No. 215.

Chen C, Shen Z. Modeling landslide susceptibility in forest-covered areas in Lin'an, China, using logistic regression, decision tree, and random forest. Remote Sensing 2023;15(18):Article No. 4378.

Chen WP, Pan L. Applying population-based evolutionary algorithms and a neuro-fuzzy system for modeling landslide susceptibility. Catena 2019;172:212-31.

Food and Agriculture Organization of the United Nations (FAO). Forests and Landslides: The Role of Trees and Forests in the Prevention of Landslides and Rehabilitation of Landslide-Affected Areas in Asia. Rome: FAO; 2013.

Harrison O. Machine learning basics with the K-nearest neighbors algorithm [Internet]. 2018 [cited 2025 Jul 15]. Available from: https://towardsdatascience.com/machine-learning-basics-with-the-k-nearest-neighbors-algorithm-6a6e71d01761.

James G, Witten D, Hastie T, Tibshirani R. An Introduction to Statistical Learning. Vol. 112. New York: Springer; 2013. p. 18.

Le TT, Kaneko S. Landslide detection analysis in north Vietnam based on satellite images and digital geographical information: Landsat 8 satellite and historical data approaches. Japan Society of Civil Engineering 2017;73(5):239-49.

Murgia I, Vitali A, Giadrossich F, Tonelli E, Baglioni L, Cohen D, et al. Effects of land cover changes on shallow landslide susceptibility using SlideforMAP software (Mt Nerone, Central Italy). Land 2024;13(10):Article No. 1575.

Pham BT, Jaafari A, Prakash I, Bui DT. Naïve bayes-random subspace hybrid model for landslide susceptibility mapping. Geoscience Frontiers 2021;12(1):101-12.

Schmaltz EM, Steger S, Glade T. The influence of forest cover on landslide occurrence explored with spatio-temporal information. Geomorphology 2017;285:250-64.

Thai Nguyen Forest Protection Department (TNFPD). Forest Inventory Data 2023. Thai Nguyen, Vietnam: TNFPD; 2024.

Tran T, Trieu H, Bantayan N. GIS-based hybrid machine learning for landslide susceptibility assessment in Thai Nguyen Province, Vietnam. Proceedings of the 2024 Geoinformatics for Spatial-Infrastructure Development in Earth and Allied Sciences (GIS-IDEAS) Conference IEEE Special Track; 2024 Dec 11-13; Grand Vista Hotel, Chiang Rai: Thailand; 2024.

Vietnam Administration of Forestry (VAF). Vietnam Annual Forest Data. Hanoi: VAF; 2024.

Vietnam Disaster Management Authority (VNDMA). Vietnam 20-year Floods and Landslides Handbook. Hanoi: VNDMA; 2023a.

Vietnam Disaster Management Authority (VNDMA). Disaster Management Report. Hanoi: VNDMA; 2023b.

Vietnam General Statistics Office (VNGSO). Statistical Yearbook of Vietnam 2023. Hanoi: Statistical Publishing House; 2024.

Wang Y, Yin K, Zhang N, Zhang L. Landslide susceptibility mapping using recurrent neural network frameworks. Landslides 2022;19:437-52.

World Bank (WB). Country Forest Note: Vietnam. Washington, DC: WB; 2019.