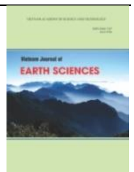
Vietnam Journal of Earth Sciences, 47(3), 430–446, https://doi.org/10.15625/2615-9783/23462

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



Integrating remote sensing and artificial intelligence for landslide detection and susceptibility analysis along tourism routes in Da Bac district, Hoa Binh province, Vietnam

Kinh Bac Dang[1], Thi Thu Huong Hoang[1,*], Hieu Nguyen[1], Kim Chi Vu[2], Tuan Linh Giang[2], Damien Closson[3], Thi Dieu Linh Nguyen[1], Thi Ngan Do[1]

[1]*VNU University of Science, Vietnam National University, Thanh Xuan, Hanoi, Vietnam*
[2]*VNU Institute of Vietnamese Studies and Development Science (VNU-IVIDES), Vietnam National University, Thanh Xuan, Hanoi, Vietnam*
[3]*Ministry of Defense, Belgium*

Received 23 May 2025; Received in revised form 05 September 2025; Accepted 17 September 2025

ABSTRACT

The occurrence of natural disasters, especially with landslides, threatens mountainous districts and has serious consequences on local tourism development. Future disaster management must develop efficient innovative tools to control the rising frequency and intensity of landslides due to the impacts of economic development and climate change. Minimizing the risk and effects of these occurrences relies on the establishment of an optimal early warning system. This study focuses on the integration of artificial intelligence approaches to identify landslides and evaluate their susceptibility, with an emphasis on early warning systems on tourist routes in Da Bac district. As the first tool in the system, advanced deep learning models using satellite data at high resolution assist in identifying landslides. As a result, a developed DeepLab-v3 model demonstrated high performance by reaching 0.213 dice coefficient and 96.8% accuracy for landslide detection without restrictions from specific input resolution sizes. As the second tool, various machine learning tools, such as Random Forest and Support Vector Machine, utilize the identified landslide locations from the first tool to assess and map their susceptibility based on environmental and human-made factors. Accordingly, the study proposed an early warning system for landslide disaster management using real-time ecological factors and historical data. The proposed integrated system helps tourists and local communities take preventive actions that reduce landslide impacts, thus achieving safety goals in tourism activities, particularly in the Da Bac district of Hoa Binh province, Vietnam. It enhances strategies to minimize risk, increases the ability to predict landslide-prone tourist areas, and aids in implementing sustainable tourism in the future.

Keywords: Landslide warning system, machine learning, deep learning, mountainous tourism area, Vietnam.

1. Introduction

According to the United Nations World Tourism Organization (UNWTO), records indicated that international travel reached 1.4 billion people in 2018 for outdoor activities in mountainous landscapes, including hiking and skiing (Lee and Jayakumar, 2021; Leuven, 2014). However, tourists have a high risk of injury from landslides due to unfamiliarity with the terrain and a lack of knowledge about warning signs (Luu et al., 2023; Newsome

*Corresponding author, Email: huonghtt@hus.edu.vn

and Dowling, 2018). For example, Annual landslides leave tourist groups isolated for days in Vietnam's northern mountainous provinces, such as Ha Giang, Lao Cai, and Hoa Binh provinces (Duc et al., 2023). At least eight people died after heavy rain led to landslides in areas in Southeast Asia following the Tropical Storm Kajiki in August 2025. There was one death in a landslide in the northern city of Chiang Mai in Thailand. The Thai Department of Disaster Prevention and Mitigation reported that several others suffered injuries in flash flooding and landslides in the northern part of the country. World Bank indicated that landslides killed thousands of lives and cause billions of dollars in property damage in 2020 (World Bank Group, 2021, 2020). The implementation of warning systems for live slope monitoring and rainfall pattern surveillance should cover the entire length of tourism pathways (Valchev et al., 2017). Emergency measures for authority intervention are needed for the detection of upcoming landslides to establish safety protections for tourists (Kubalíková et al., 2021). Landslide warnings in Nepal offer 80% better protection by combining rapid emergency rescue operations with immediate safety alerts (Gallo and Lavé, 2014). The assessment of landslide risks allows the construction of safety-oriented tourist infrastructure throughout different regions (Froude and Petley, 2018). Tourist land-use planning in the rocky areas minimizes landslide dangers by half (Tricia et al., 2019). The safety measure includes both the prevention of walking trails in dangerous regions and restrictions on building projects in vulnerable areas (Aji et al., 2021). Therefore, the detection of landslides and their associated danger must be utilized to make effective choices about developing tourism initiatives in mountain environments.

Traditionally, geologists can create detailed maps that predict landslide probability by analyzing digitized data (Duc et al., 2023; Pham et al., 2023). Based on their susceptibility, specific thresholds proposed by experts, such as slope angles or cumulative rainfall over a period, can trigger warnings or bans, thus safeguarding tourists (Aprina et al., 2024; Doan et al., 2024). Historical documents, along with geological information, have traditionally served as a valuable resource for researchers in detecting landslides (Nguyen et al., 2011; Roa-Lobo, 2007). The field of landslide identification has undergone significant advancements due to machine learning (ML) and, more specifically, deep learning (DL) advancements (Naveen et al., 2022). DL models deliver much more effective landslide detection and vulnerability assessment capabilities compared to traditional ML methods (Nhu et al., 2020). The detection of landslides using Support Vector Machines (SVMs) and Random Forests demonstrates proven effectiveness in studies of Phong et al (2020), Prakash et al (2021), and Nguyen et al (2024). Accordingly, the frequency and location of landslides can be mapped. The location of landslides detected from above models helps tourism officials observe changes in hazard information before prediction and choose an appropriate response plan.

The current landslide risk assessment techniques have been developed from ML models (Khan et al., 2021; Zhao and Lu, 2018). The ability of Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) to identify intricate associations within large datasets related to landslide factors was demonstrated by (Ghorbanzadeh et al.). The analysis of satellite imagery using CNNs becomes possible through images of confirmed landslides to achieve accurate landslide probability maps (Catani, 2021). A combination of current weather monitoring data and previous

[1]https://www.euronews.com/2025/08/27/

precipitation records enables them to calculate landslide probabilities by considering changing environmental conditions. For instance, a CNN can detect subtle changes in a hill, like increasing moisture levels or minor movements, from years of satellite data, enabling it to forecast a landslide before it occurs (Prakash et al., 2021). RNNs are highly effective with sequential data, such as rainfall patterns (Alom et al., 2019). A forecasting ability enables authorities to discover potential dangers before they issue critical safety warnings during threatening situations (Casagli et al., 2023). The mentioned studies have proven that modern technological tools based on ML and DL systems enhance both landslide prediction analysis and detection capabilities, which strengthens mountain tourism protection security (Tofani et al., 2013). However, to date, this integrated system has not been developed or implemented in any country.

This work aims to propose a DL model that can identify signs of landslides using WorldView-2 satellite images and an ML model that can evaluate the landslide hazards by analyzing relevant environmental and socio-economic aspects. The proposed integrated system utilizes various methods to connect field observations with detailed remote sensing data that identifies landslides in WorldView-2 datasets, as explained in Section 2.3. Sections 2.4 and 2.5 present the advancements in different machine-learning techniques used in assessing landslide risks. Testing of the outcome models was conducted in Da Bac, Hoa Binh Province, a designated tourism area in Vietnam, by combining diverse data types with artificial intelligence processing methods.

2. Materials and methods

2.1. Case study

The case study in Da Bac, Hoa Binh province, Vietnam (Fig. 1), was chosen to test the proposed artificial intelligence system to monitor landslide traces and analyze landslide hazards along tourism routes. Based on the potential of tourism development, in recent years, the Da Bac district has prioritized infrastructure investment, promotion, and the implementation of solutions to foster tourism development, thereby establishing appealing tourist destinations (Hung et al., 2015). Notably, managers emphasised investments in transportation infrastructure, including the expansion of the Hien Luong-Tien Phong route, which is anticipated to be finalized by 2025. Sung village (Cao Son commune), Ke village (Hien Luong commune), and Da Bia village (Tien Phong commune) are the district's most prominent community tourism destinations (Bui et al., 2012b). The Institute of Geosciences and Mineral Resources under the Ministry of Natural Resources and Environment of Vietnam identifies Hoa Binh provincial landslides as highly prone to danger (Doan et al., 2024). A total of 89 sites have been documented to carry a high risk of sudden landslides, which affected 4,661 households during the 2010s. The Da Bac District People's Committee has established multiple plans to anticipate various hazardous circumstances, including unpredictable landslide threats from complex terrain conditions and rainy seasons (Hung et al., 2015). Landslides have impacted 15/17 villages in the entire Da Bac district in 2019.

Landslides are particularly prevalent in regions with steep slopes, complex lithology, or roads with positive gradients during the rainy season. In particular, the section of National Highway 6 that passes Doc Cun at Km79+100, at route 443, through Hoa Binh city and Da Bac district, experienced six landslides, resulting in a total volume of soil and debris exceeding 1,600 m^3 (Hang et al., 2021). In recent years, individuals have ceased to cut down trees on the summit of the hill to protect the land. However, the risk of landslides remains constant during the rainy season. The majority of landslides occurred at night, which renders them even more dangerous. Therefore, local authorities need a comprehensive warning system to help minimize unexpected incidents that may happen to residents and tourists in the future.

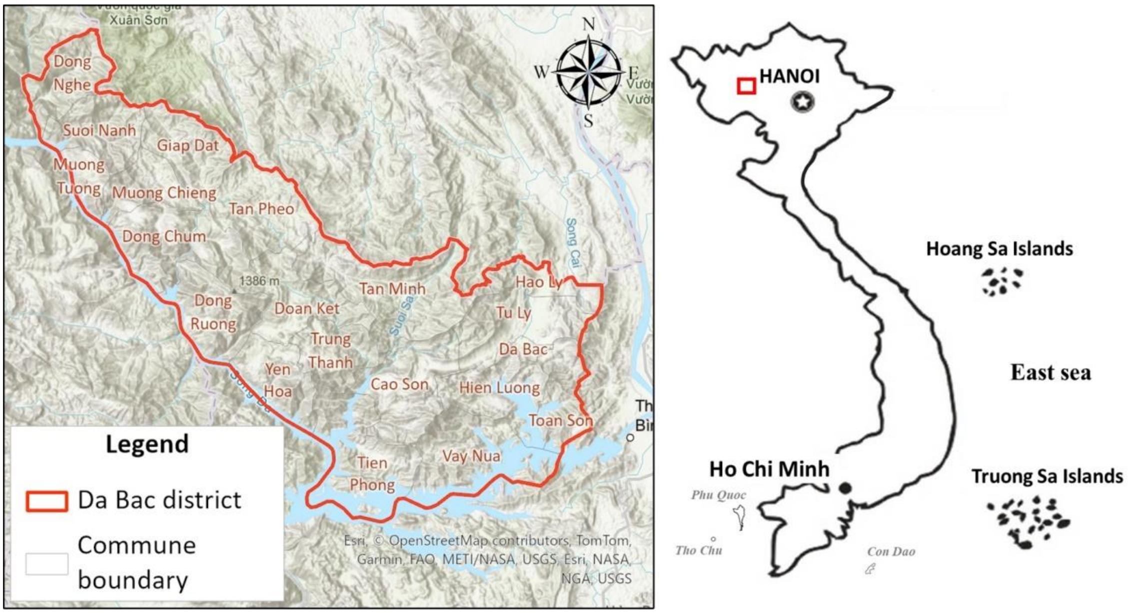


Figure 1. Location of research area

2.2. Research process

The research process is divided into three steps and shown in Fig. 2. Step 1 uses a DL model developed from a previous study by the research team using Sentinel-2 remote sensing to identify landslide locations in the past. The content of this step was summarized in section 2.2. Because the results of the DL model are new, they are presented in Section 3.1. Next, step 2 of the process involves developing an ML model using geospatial information collected from past landslide locations in step 1. The geospatial data and landslide factors will be filtered to select the most critical factors for developing the ML model. After selecting the most optimal ML model, it was applied to predict landslides for 12 consecutive months, thereby identifying traffic areas with the highest landslide risk.

2.3. Step 1: Application of a deep learning model for landslide detection

The application of DL models to identify landslide traces in satellite images was presented in the paper titled *"Deep learning models integrating multi-sensor and temporal remote sensing to monitor landslide traces in Vietnam"* of our research group, published in the "International Journal of Disaster Risk Reduction" journal in 2024. Once the DL model was successfully developed, the authors utilized it to detect landslides in WorldView-2 satellite images, eliminating the need for additional training data. In this study, the trained DL model was applied for landslide detection at Da Bac, Hoa Binh province, Vietnam. The description to develop this model is briefly explained in Appendix B. The outcome of this model was shown in Section 3.1.

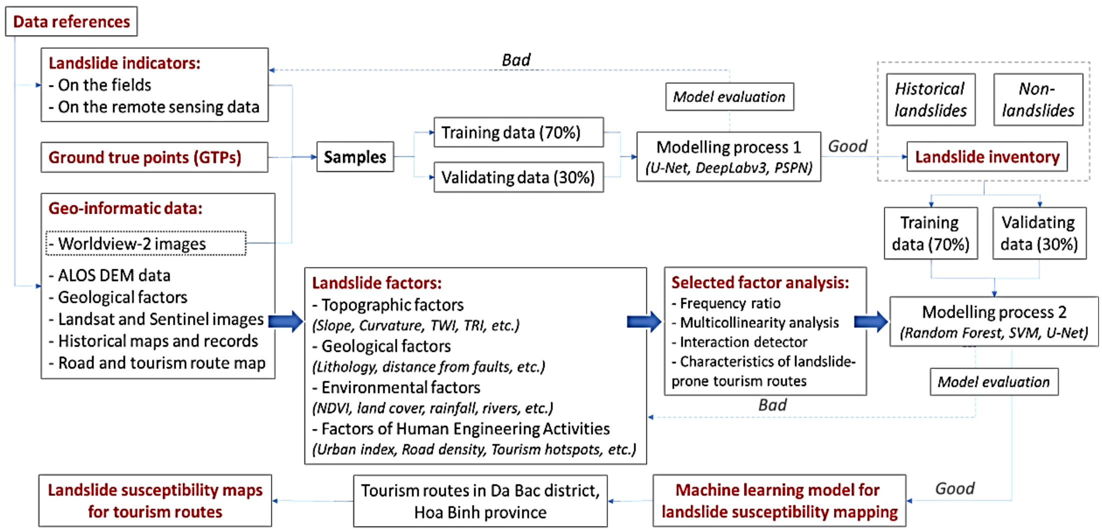


Figure 2. Framework to detect landslide traces and assess landslide hazards based on artificial intelligence models

2.4. Step 2: Preparing landslide conditioning factors

A landslide is generated when rocks or soil, together with other sedimentary materials, descend a sloping terrain (Dang et al., 2018; Nguyen et al., 2024). These events are complex and result from various interrelated factors. Learning these factors enables more effective landslide prevention. This research reviewed expert opinions, scientific literature, and available data to provide insights into landslide hazard (Tran et al., 2024). Accordingly, landslide conditioning factors include geological, lithological, topographical, hydrological, etc., and social ones (Geertsema et al., 2009; Khan et al., 2021). However, some of them can have a similar meaning or interaction effects with each other. Therefore, once all data were collected, the correlation between them needs to be analyzed to eliminate less meaningful factors in assessing landslide hazard. The description of the considered factors is seen in Fig. 3 and explained in detail in Appendix C.

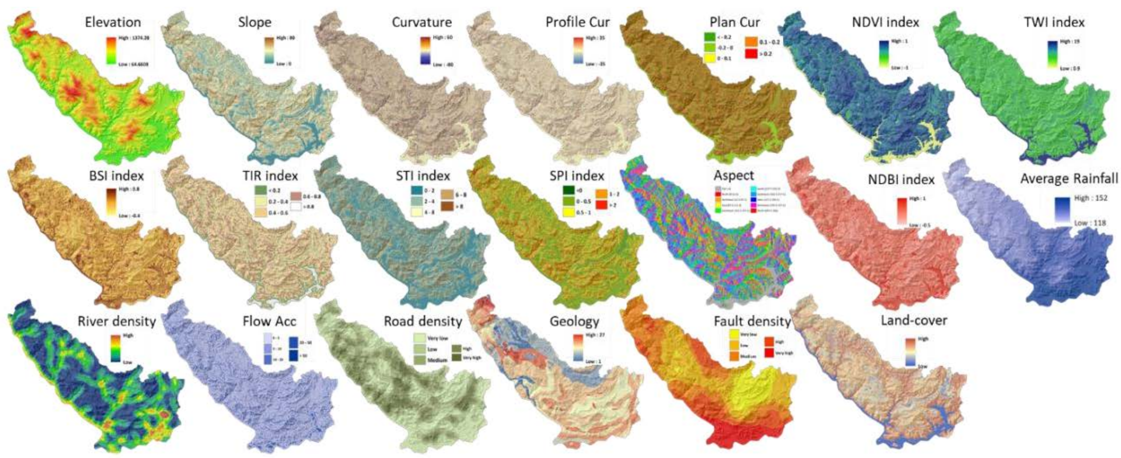


Figure 3. Maps of landslide conditioning factors

2.5. Step 3: Selecting architectures for the ML model for landslide susceptibility mapping

Traditional landslide hazard mapping by interdisciplinary experts requires substantial human effort and subjective evaluations, yet it lacks suitability when handling advanced geographic data (Schweigl and Hervas, 2009; Zhao and Lu, 2018). ML technology presents an effective and reliable response through its operational solution. For example, the Support Vector Machine model masters pattern discovery when processing extensive data collections that help determine landslide zones by evaluating slope data and plant cover patterns as well as past landslide sites (Abdo and Richi, 2024). Meanwhile, Random Forests achieve ensemble learning through multiple decision trees that analyze random subsets of landslide features before identifying them (Sharma et al., 2024). The analysis of large datasets by ML algorithms helps detect elaborate associations that lead to landslide risk factors. The algorithms detect relationship patterns between predictor variables and past landslide events to forecast existing landslide risks. In this study, the authors employed five advanced ML algorithms Random Forest, SVM, KNN, LDA, and Decision Tree to map landslide susceptibility (Agboola et al., 2024; Sun et al., 2024). All values of landslide conditioning factors (mentioned in Section 2.3) were extracted for each "landslide" and "non-landslide" sample (mentioned in Section 2.2). The input data for training ML models is a matrix of landslide conditioning factors in these samples. They were separated into two parts, including a training and validation group (with a ratio of 80:20), before being added to the training model process. The details of each model are presented in Appendix D.

Regarding model performance, accurate evaluation of ML models supports effective landslide susceptibility forecasting and prevents both model overfitting and underfitting. The performance assessment of trained models depends on loss function values and different parameters such as ACC, Kappa, Sensitivity, Specificity, Positive Predicted value, Negative Predicted value, and the area under the curve (AUC) to evaluate the precision between prediction and training validation labels (Abdo and Richi, 2024; Pham et al., 2022). The description of each measurement value was explained in detail in Appendix E.

2.6. Scenario development for monthly landslide susceptibility

After choosing the best model, the model can be used to assess and map landslide susceptibility for new regions. Although the input data consists only of sample points in Da Bac district, Hoa Binh province, Vietnam, the landslide susceptibility in the whole district has been mapped, especially along tourism routes. Once the ML model is completed, no more sample is needed. The input data in scenario development are only maps of landslide conditioning factors. The landslide susceptibility map can be updated based on precipitation. Other maps related to the eight input variables were maintained, while the precipitation map can be changed to visualize the changes in landslide susceptibility every month. The average value of precipitation was also selected to observe the background of landslide susceptibility in general. Additionally, the authors selected the precipitation maps for 12 months as one main input for running the ML model. Accordingly, the changes of landslide susceptibility along the tourism routes were estimated in twelve months of a year in the Da Bac, Hoa Binh province.

3. Results

3.1. Landslide traces detected based on a deep learning model

Based on the Deeplab-v3 model with

Densenet169 architecture and 256×256 input data, 3590 landslide traces were found in Da Bac district (Fig. 4 - left). The identified traces are not only located along roads but also in remote areas or on hillsides. The high-density areas could be found in the Northwest region of Da Bac district, observed in satellite images from 2018 and 2019. It coincides with the rainy season in the entire Northwest region of Vietnam. In particular, landslide traces with high density are located along streams and the eastern part of the Da River valley.

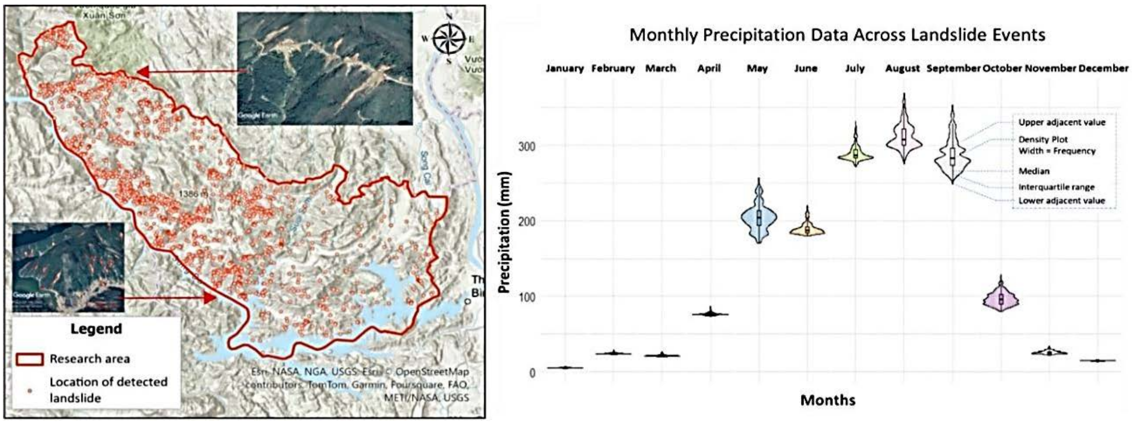


Figure 4. Location of landslide traces detected from deep learning models (left) and their allocation in each month of a year (right) in Da Bac, Hoa Binh province, Vietnam

Meanwhile, landslides around the Song Da hydroelectric dam area have a lower density, accounting for about 10%. It is also a densely populated place in the research area. Figure 4 (right) shows the density of landslide traces by month. Accordingly, landslide traces were numerous in many main tourism corridors linking communes with ecotourism locations within Da Bac, which visitors frequent to explore homestays and natural features. Various landslides were found in trekking routes in the southern part of the research area. The landslides along these roads not only pose a threat to the safety of transport but also directly influence the stability of the community-based tourism activities, which depend so much on the availability of these roads. Thus, mapping landslide-prone areas along tourist routes would be vital for hazard management and sustainable tourism development within the district. This indicates that the number of landslide traces increases from May to October, particularly in July, August, and September. Rainfall during these months reaching above 150–170 mm can be considered the threshold for landslides here. In October, as the storms retreated south of Vietnam, the number of landslides in the area also decreased.

3.2. Calibration and verification of machine learning for landslide susceptibility mapping

Through analyzing the correlation between 20 variables affecting landslide hazards, variables with poor correlation or interactive effects with other variables were eliminated. This elimination work is based on the analysis of the AIC and BIC values. Thereby, the input variables can be separated into three groups of variables for the ML model development (Appendix A1). The first group has all 19 input variables. The second group includes 14 variables after filtering out variables with interaction effects. The third group filters out variables with a correlation lower than 0.1.

The nine key variables included will include BSI, NDVI, NDBI, TIR, flow accumulation, Slope, DEM, distance to streams, faults, and roads. The correlation between this variable and the possibility of landslides is presented in Appendix A2. Accordingly, the variation of the variables is limited and not entirely linear. Some relatively linear negative correlations appear in the variables "distance to rivers and streams" and "distance to roads". Landslide traces are found mainly near rivers, streams, and roads, increasing the probability of finding landslides in these areas. Appendix A3 also shows the warning threshold for landslides at other variables such as BSI, NDVI, Slope, etc. This shows that applying linear models in landslide risk assessment can face many difficulties.

The results of developing 27 ML models in landslide risk assessment are presented in Table 1. Accordingly, most models have ACC and AUC values above 80%.

Table 1. Model performance to assess landslide susceptibility in tourism routes

No.	No. Variables	Model	ACC	Kappa	Sensitivity	Specificity	Pos Pred Value	Neg Pred Value	AUC
1	9	Logistic Regression	0.84	0.68	0.86	0.81	0.82	0.86	0.91
2	**9**	**Random Forest**	**0.90**	**0.80**	**0.90**	**0.90**	**0.90**	**0.90**	**0.96**
3	9	SVM	0.84	0.68	0.86	0.82	0.83	0.86	0.92
4	9	KNN	0.73	0.45	0.73	0.72	0.72	0.73	0.92
5	9	LDA	0.84	0.68	0.82	0.86	0.86	0.82	0.91
6	9	Decision tree	0.85	0.70	0.85	0.85	0.85	0.85	0.86
7	9	PSO-CNN-64	0.85	0.70	0.86	0.84	0.84	0.86	0.88
8	**9**	**PSO-CNN-128**	**0.88**	**0.75**	**0.89**	**0.86**	**0.87**	**0.89**	**0.88**
9	**9**	**PSO-CNN-256**	**0.88**	**0.75**	**0.88**	**0.87**	**0.88**	**0.88**	**0.88**
10	14	Logistic Regression	0.85	0.69	0.86	0.83	0.83	0.86	0.92
11	**14**	**Random Forest**	**0.91**	**0.81**	**0.89**	**0.92**	**0.91**	**0.90**	**0.96**
12	14	SVM	0.85	0.70	0.86	0.84	0.84	0.86	0.92
13	14	KNN	0.64	0.27	0.57	0.71	0.65	0.63	0.68
14	14	LDA	0.85	0.70	0.83	0.87	0.87	0.83	0.92
15	14	Decision tree	0.86	0.71	0.83	0.88	0.87	0.84	0.86
16	**14**	**PSO-CNN-64**	**0.88**	**0.76**	**0.86**	**0.90**	**0.89**	**0.87**	**0.86**
17	**14**	**PSO-CNN-128**	**0.88**	**0.76**	**0.87**	**0.89**	**0.89**	**0.87**	**0.88**
18	14	PSO-CNN-256	0.85	0.70	0.86	0.84	0.84	0.86	0.85
19	19	Logistic Regression	0.85	0.70	0.86	0.84	0.84	0.86	0.92
20	**19**	**Random Forest**	**0.90**	**0.80**	**0.89**	**0.91**	**0.90**	**0.89**	**0.96**
21	19	SVM	0.85	0.71	0.86	0.85	0.85	0.86	0.91
22	19	KNN	0.76	0.51	0.67	0.84	0.80	0.72	0.76
23	19	LDA	0.85	0.70	0.84	0.86	0.86	0.84	0.91
24	19	Decision tree	0.86	0.72	0.86	0.85	0.85	0.86	0.87
25	19	PSO-CNN-64	0.86	0.73	0.92	0.81	0.83	0.91	0.86
26	**19**	**PSO-CNN-128**	**0.88**	**0.76**	**0.85**	**0.90**	**0.90**	**0.86**	**0.88**
27	**19**	**PSO-CNN-256**	**0.88**	**0.76**	**0.90**	**0.86**	**0.87**	**0.90**	**0.88**

However, the Kappa index of the models has apparent differences. The study indicated nine models with a Kappa accuracy over 0.75. In these models, three of them use nine variables, three models use 14 variables, and three models use 19 variables. The models with the lowest performance utilize the KNN structure, as this structure is unable to reproduce the non-linearity between variables in assessing landslide hazard. The ML structural models used achieve high performance, including RF and PSO-CNN. In which the RF model has a Kappa index greater than 0.8 in all cases, and other indexes are approximately 0.9. It is also shown through the ROC curve in Appendix A3.

Finally, the RF model utilizes nine selected input variables to construct a landslide susceptibility map for subsequent steps. The selection of the RF model is specifically appropriate in the sphere of tourism management since it guarantees both stability and the simplicity of results interpretation. With very few input variables and yet high accuracy, this model can help managers establish, within a minimal time frame, the areas where there is a risk of landslides along the main tourist routes and make the right decision to implement the necessary safety measures, without the need to work with complex data sets. RF is the best instrument to incorporate landslide risk assessment in sustainable tourism development planning due to the compromise between reliability and practicality.

3.3 Monthly landslide susceptibility maps

After the ML model was completed, all data on nine variables along tourist routes in the Da Bac area, Hoa Binh province, were included in the model to evaluate landslide susceptibility. With an assessment based on average rainfall, the high potential of the landslide susceptibility is mainly concentrated in the western region, specifically along inter-district and inter-commune roads, and areas with experiential tourist routes related to mountain climbing and exploration travel. The proportion of areas with a landslide probability of over 50% accounts for more than one-third of the study area. The central part of the district (the area between points 1, 2, and 3 in Fig. 5) has an average susceptibility level, indicating suitability for residential development. Regarding the landslide points located along the roads, the field survey revealed that the government has implemented reinforcements to ensure the safety of people traveling. However, landslides still occur during the rainy season (Sub-Figs. 1, 2, and 3 in Fig. 5). For tourist routes spontaneously created by people, as shown in Fig. 5, the treatment and reinforcement of these routes are difficult due to accessibility to the disaster area. It can be observed throughout the western tourist route and the outer eastern edge of the residential area.

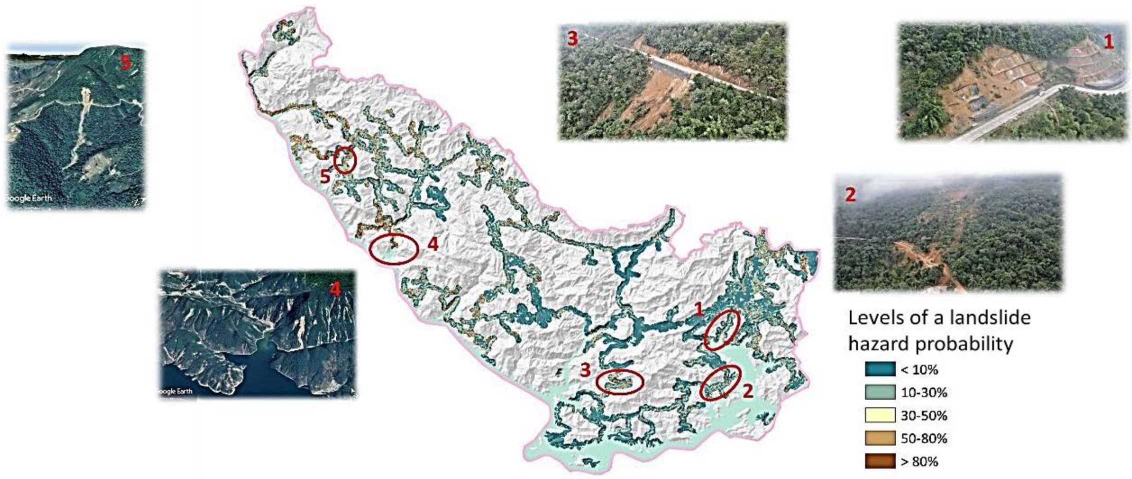


Figure 5. Landslide susceptibility in tourism routines based on average precipitation

Through assessing landslide susceptibility in each month of the year, the probability of this hazard occurring along tourist routes is low before April (Fig. 6). In January, February, and December, the rainfall in this district is usually less than 50 mm, making the probability of landslides very low. In March and November, rainfall is about 50–100 mm, and landslide susceptibility increases in some areas, but in small parts with low probability. From May to October, especially in July, August, and September, landslides occur frequently along tourist routes. Rainfall at this time ranges from 150 to 350 mm. Landslide warning zones are only scattered at a few points in the west, along the Da River valley. Most points have a landslide probability of over 80%. This can be easily seen in Fig. 7. On average, 15–17% of tourist routes in the area are affected by landslides. In particular, the months when landslides occur are a time for tourism in the locality and in Vietnam.

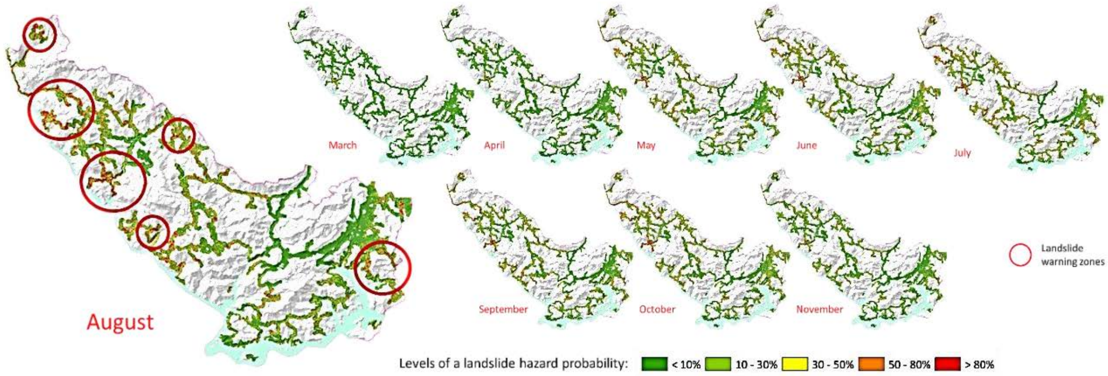


Figure 6. Landslide susceptibility in tourism routines based on monthly precipitation in Da Bac, Hoa Binh province, Vietnam

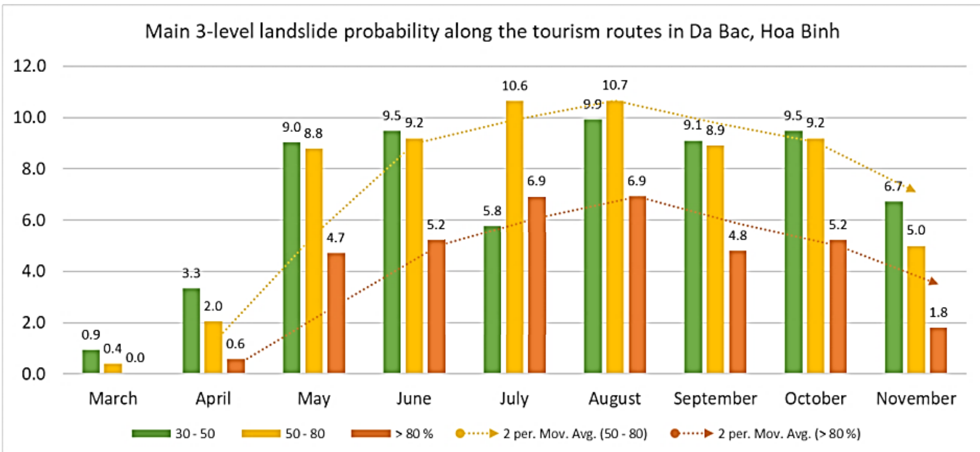


Figure 7. Landslide susceptibility levels in tourism routines in each month in Da Bac, Hoa Binh province, Vietnam

4. Discussions

4.1. Integrated landslide warning system in tourism routes

Based on the development of two artificial intelligence (AI) models to identify landslide traces and assess landslide susceptibility, this study proposes the development of an integrated AI landslide warning system. Importantly, timely landslide warnings require a comprehensive and cohesive approach, allowing for the prediction of landslide hazards every month. The increase in precipitation and the number of rainy days significantly changes the probability of landslide hazards, as seen in Figs. 6 and 7. To get high levels of precision and productivity, this system must include geographical, climatic, and human-made factors, with AI technology.

In the integrated system, the first step involves acquiring high-resolution satellite imagery data from platforms such as Google Earth Pro and Worldview-2, as done in Section 2.2. The data has to be evaluated and partitioned into smaller segments for the purpose of training the model. Data processing involves the identification of landslide indicators via the use of field observations and satellite data to generate dependable and practical datasets. To achieve a high level of accuracy in detecting landslides, it is necessary to train the model using processed data and techniques such as image segmentation and data labeling, as done in this research and other studies, including Cheng et al. (2013) and Naveen et al. (2022). Different advanced DL models like DeepLab-v3 or U-Net with the ResNet-34 framework proved their potential to identify landslide traces on high-resolution satellite images in this study.

Once the landslide traces were identified, the computer must evaluate related factors, including geology, lithology, topography, hydrology, and artificial effects, to determine the state of ancient landslide tracks. Assess landslide risk by using indicators such as STI, TRI, TWI, and SPI as done by Tien Bui et al. (2019). This facilitates the identification of high-risk areas and acts as input for ML systems. The development of landslide probability assessment models can be achieved through SVM, RF, and Decision Tree algorithms, as noted by Bui et al. (2012). The model requires processing of data through conditional factor analysis to achieve accurate predictions during its training phase, along with assessments. The AI system performs landslide detection functions by integrating prediction models within its warning system to anticipate impending landslide hazards. The system incorporates measurements of temperature alongside precipitation data to establish predictions about forthcoming landslide dangers, which it follows persistently.

In the context of Da Bac district and similar mountainous areas, 17% of the detected landslide traces are located along main transportation and tourism routes. It posed a direct threat to road safety and created disruptions for visitors traveling to ecotourism villages, homestays, and cultural sites. As such, it is essential to incorporate road-network data into the AI-based warning system to prioritize hazard alerts on parts of the road that tourists use most often. In this way, the system can not only help mitigate disaster risks but also enhance the resilience of tourism infrastructure and provide safer access to destinations. The operating warning systems should establish immediate capabilities to exchange information with regulatory bodies and local communities for proper preparation. Accordingly, this study took on the challenging task of building an artificial intelligence system that links landslide identification to warning systems for potential tourism benefits. AI integrates with landslide condition factor analysis through predictive warning systems to provide opportunities for lower landslide risks and reduced loss potential. It not only facilitates natural catastrophe mitigation in time, but it also fosters the sustainable tourism development of areas susceptible to landslides.

4.2. Monthly landslide warning for local tourism routes in Hoa Binh province

The case study is a key landslide area in

Hoa Binh province, which affects the tourism development of the district. Its most significant impacts are on infrastructure, such as road damage, or on the tourism industry by lowering tourism numbers due to fear of landslides. For example, Ban Lac village was severely affected, with tourism capacity in 2023 at a mere 60–70% and in 2024 at a devastatingly low 10%. This study has pointed out landslide hotspots over time and space. Local authorities need to have warnings and solutions to minimize landslides, thereby contributing to safe tourism. In terms of time, the peak months for landslides are in the rainy season (May to October), especially in July, August, and September. Landslides become exceptionally rare between November and April of the upcoming year during the dry season period.

The study displays very comparable findings to the research performed by Hoang and Nguyen (2022) on the tourism climate index (TCI). The research data indicate that tourism safety levels vary throughout the year, with the first four months offering ideal weather conditions for tourism activities and stable Lake Hoa Binh water levels, accompanied by minimal geological movements, resulting in secure tourism conditions. This is also the springtime with many festivals, so many visitors are coming to Hoa Binh. The period from May to June marks the beginning of the rainy season. At this time, tourism activities can still take place, but attention should be paid to early flooding. From July to October is the peak of the rainy and flood season, with warnings about safety and limiting tourism activities. The end of the year, from November to December, also features pleasant weather, with few landslides, making it a safe environment for tourism activities. The risk map can be applied to seasonal activities in the context of tourism management. Speaking of which, in the months (particularly in July-September) when the weather is at its wettest, adventure tourism like trekking or mountain climbing should be restrained; instead, cultural tourism, festivals, or water tourism to Hoa Binh Lake should be promoted to be safe. Ecotourism, community experiences, and trekking can also be extended during the dry months due to the low risk of landslides. In terms of space, the study showed that landslide warning zones are only scattered at a few points in the west, along the Da River valley. Most routes have a landslide probability of over 80%. Such dangerous locations occupy most of the critical road networks, like provincial roads linking Hoa Binh city with Da Bac and other upland communes that are the main road networks serving community-based tourism and ecotourism sites. The local transport is not only disrupted by landslides on these roads, but also poses serious hazards to the tourists visiting homestays, cultural villages, and natural attractions. Therefore, landslide risk information must be considered in the planning of tourism routes to provide a safe environment to visitors and ensure the sustainability of road-based tourism activities. To avoid these landslide points, tourists can travel to Da Bac by waterway on Hoa Binh Lake instead of by road, ensuring their safety.

4.3. Further database for the early warning landslide system

Discovering evidence of landslides requires a comprehensive understanding and high-quality data. This study utilized multi-temporal and multi-spatial satellite imagery from different source platforms, as proposed by Mohan et al. (2021) and Ghorbanzadeh et al. (2022). High-resolution satellite imagery is essential since the majority of landslide traces are subtle and difficult to detect. The high resolution of these data enables the detection of subtle characteristics, such as topographic irregularities and structural fractures, caused by changes in plant distribution. Identifying landslides depends on these critical signs found in the geological environment. The processing method for DL models uses dimension reduction on images by generating 64×64, 128×128, and 256×256 pixel parts. Researches thus far has failed to implement this process. The sub-images function as tools to reduce random data while improving the detection of fine elements in images.

Field surveys are essential for validating data accuracy, since they ensure that digital information is not limited to images but accurately reflects the actual conditions on the ground. This approach provides the authenticity of landslides exhibited on satellite images by cross-referencing them with actual conditions observed at researched landslide locations. Training DL models with high-quality information ensures data reliability according to this method. Continued modifications in sub-image dimensions, ranging from 64×64 to 128×128 and 256×256 pixels, help the DL model detect landslides at different scales, from minor to extensive features. Small subsampled images demonstrate apparent minor landslide indications, while larger subsampled images present the best capabilities for showing minor landslide signs.

The landslide susceptibility assessment demands data collection about geology, together with soil composition and topographic and hydrological conditions, alongside human activities and geological characteristics. Each research project selects its own group of variables depending on its focus area, which leads to different variables ranging from six by Agboola et al. (2024) up to 13 by Nguyen et al. (2024) and 15 by Tien Bui et al. (2019). The evaluation process requires precise selection and complete examination of multiple potentially predictive variables that aim to display landslide potentials with accuracy. Slope, geological, and hydrological features are significant determinants of the likelihood of landslides occurring in an area. On the other hand, the density of plants influences the stability of the soil. The correlations among the components allow eliminating some variables with the least significant or highly correlated features, therefore, controlling the data set and focusing on the most relevant variables. Other studies tend to omit this step, such as Sharma et al. (2024) and Pham et al. (2022). The PCA analysis proved that some variables, such as proximity to roads and rivers, and precipitation, have a substantial impact on the likelihood of landslides. The ML techniques in this study analyzed all these factors to accurately simulate the complex relationships between them and with landslide occurrences. This database enables mapping by providing an overall representation of the factors that drive landslides, as well as by enabling the accurate location of high-risk zones through ML techniques. In addition, the structured database can be shared with the tourism sector to assist in risk communication and safe route planning. Integration of tourism road and destination maps and hazard data can enable tourism operators to develop contingency strategies and provide visitors with safer tourism choices. This type of cross-sector data sharing contributes to improving the interconnection of sustainable tourism development and disaster risk reduction.

5. Conclusions

The paper proposes an AI-based early warning system to predict landslide danger. Firstly, DL detection enhanced detection accuracy, making the models suitable for landslide trace monitoring. AI-based landslide detection employs high-resolution satellite images and DeepLab-v3 algorithms. Secondly, an AI-based model can be used to assess landslide susceptibilities based on geological, topographic, hydrological, and anthropological variables, utilizing a Random Forest algorithm for assistance. By forecasting monthly accidents, authorities can develop strategic land-use plans, build infrastructure, and create evacuation plans for sensitive sites. The proposed system proved its optimal performance in this study. According to the system, tourists in the northern part of Vietnam can travel from November to April, which is doable in May and June, while frequent rains limit visitation from July to October. Many of the detected landslide-prone areas coincide with main road corridors that connect tourism destinations, homestay villages, and cultural sites. Warnings along these routes are critical because landslides not only disrupt traffic but also directly endanger tourists traveling by road to remote attractions. Integrating route-specific hazard information into the early warning system ensures that tourism stakeholders and visitors can adjust travel schedules, select alternative routes, or shift to safer modes of transport. Accordingly, the application of the whole proposed system in landslide detection and susceptibility assessment improves disaster management and makes communities in natural hazard zones safer.

Funding

This research is funded by the Vietnam National Foundation for Science and Technology Development (NAFOSTED) under grant number NCUD.05-2022.04

References

Abdo H.G., Richi S.M., 2024. Application of machine learning in the assessment of landslide susceptibility: A case study of mountainous eastern Mediterranean region, Syria. J. King Saud Univ. Sci., 36, 103174. https://doi.org/10.1016/j.jksus.2024.103174.

Agboola G., Beni L.H., Elbayoumi T., Thompson G., 2024. Optimizing landslide susceptibility mapping using machine learning and geospatial techniques. Ecol. Inform., 81, 102583. https://doi.org/10.1016/j.ecoinf.2024.102583.

Aji R.R., Faniza V., Tarlani Damayanti V., 2021. Landslide Disaster Engineering in Tourism Potential Area. IOP Conf. Ser. Earth Environ. Sci., 830. https://doi.org/10.1088/1755-1315/830/1/012036.

Alom M.Z., Taha T.M., Yakopcic C., Westberg S., Sidike P., Nasrin M.S., Hasan M., Van Essen B.C., Awwal A.A.S., Asari V.K., 2019. A state-of-the-art survey on deep learning theory and architectures. Electron., 8, 1–67. https://doi.org/10.3390/electronics8030292.

Aprina P.U., Santoso D., Alawiyah S., Prasetyo N., Ibrahim K., 2024. Delineating geological structure utilizing integration of remote sensing and gravity data: a study from Halmahera, North Molucca, Indonesia. Vietnam Journal of Earth Sciences, 46(2), 147–168. https://doi.org/10.15625/2615-9783/20010.

Bui T.D., Hoang N.D., Nguyen H., Tran X.L., 2019. Spatial prediction of shallow landslide using Bat algorithm optimized machine learning approach: A case study in Lang Son Province, Vietnam. Adv. Eng. Informatics, 42, 100978. https://doi.org/10.1016/j.aei.2019.100978.

Bui T.D., Pradhan B., Lofman O., Revhaug I., 2012a. Landslide susceptibility assessment in Vietnam using support vector machines, decision tree, and naive bayes models. Mathematical Problems in Engineering, 974638, 1–26. https://doi.org/10.1155/2012/974638.

Bui T.D., Pradhan B., Lofman O., Revhaug I., Dick O.B., 2012b. Landslide susceptibility assessment in the Hoa Binh province of Vietnam: A comparison of the Levenberg-Marquardt and Bayesian regularized neural networks. Geomorphology, 171–172, 12–29. https://doi.org/10.1016/j.geomorph.2012.04.023.

Casagli N., Intrieri E., Tofani V., Gigli G., Raspini F., 2023. Landslide detection, monitoring and prediction with remote-sensing techniques. Nat. Rev. Earth Environ., 4, 51–64. https://doi.org/10.1038/s43017-022-00373-x.

Catani, F., 2021. Landslide detection by deep learning of non-nadiral and crowdsourced optical images. Landslides, 18, 1025–1044. https://doi.org/10.1007/s10346-020-01513-4.

Chen L.C., Papandreou G., Kokkinos I., Murphy K., Yuille A.L., 2018. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. IEEE Trans. Pattern Anal. Mach. Intell., 40, 834–848. https://doi.org/10.1109/TPAMI.2017.2699184.

Cheng G., Guo L., Zhao T., Han J., Li H., Fang J., 2013. Automatic landslide detection from remote-sensing imagery using a scene classification method based on boVW and pLSA. Int. J. Remote Sens., 34, 45–59. https://doi.org/10.1080/01431161.2012.705443.

Dang K.B., Burkhard B., Müller F., Dang V.B., 2018. Modelling and mapping natural hazard regulating ecosystem services in Sapa, Lao Cai province, Vietnam. Paddy Water Environ., 16, 767–781. https://doi.org/10.1007/s10333-018-0667-6.

Dang K.B., Giang T.L., Dang V.B., Phan T.T., Truong Q.H., Ngo V.L., Do T.H., Dang N.V., Forino G., 2024a. Deep learning models integrating multi-sensor and -temporal remote sensing to monitor landslide traces in Vietnam. Int. J. Disaster Risk Reduct., 105, 104391. https://doi.org/10.1016/j.ijdrr.2024.104391.

Dang K.B., Ha T., Nguyen T., Nguyen H.D., Truong Q.H., Vu T.P., 2022. U-shaped deep-learning models for island ecosystem type classification , a case study in Con Dao Island of Vietnam. One Ecosyst., 7, 23. https://doi.org/10.3897/oneeco.7.e79160.

Dang K.B., Nguyen C.Q., Tran Q.C., Nguyen H., Nguyen T.T., Nguyen D.A., Tran T.H., Bui P.T., Giang T.L., Nguyen D.A., Lenh T.A., Ngo V.L., Yasir M., Nguyen T.T., Ngo H.H., 2024b. Comparison between U-shaped structural deep learning models to detect landslide traces. Sci. Total Environ., 912, 169113. https://doi.org/10.1016/j.scitotenv.2023.169113.

Doan V.L., Nguyen B.Q.V., Nguyen C.C., Nguyen C.T., 2024. Research assessment landslide and sedimentation of Hoa Binh hydropower reservoir. Vietnam J. Earth Sci., 46(2), 203–221. https://doi.org/10.15625/2615-9783/20065.

Duc Dao Minh, Minh V.C., Yen H.H., Loc N.T., Duc Do Minh, 2023. Analysis of landslide kinematics integrating weather and geotechnical monitoring data at Tan Son slow moving landslide in Ha Giang province. Vietnam J. Earth Sci., 45(2), 131–146.

Froude M.J., Petley D.N., 2018. Global fatal landslide occurrence from 2004 to 2016. Nat. Hazards Earth Syst. Sci., 18, 2161–2181. https://doi.org/10.5194/nhess-18-2161-2018.

Gallo F., Lavé J., 2014. Evolution of a large landslide in the High Himalaya of central Nepal during the last half-century. Geomorphology, 223, 20–32. https://doi.org/10.1016/j.geomorph.2014.06.021.

Geertsema M., Highland L., Vaugeouis L., 2009. Environmental Impact of Landslides, in: Sassa, K., Canuti, P. (Eds.), Landslides - Disaster Risk Reduction. Springer Berlin Heidelberg, Berlin, Heidelberg, 589–607. https://doi.org/10.1007/978-3-540-69970-5_31.

Ghorbanzadeh O., Shahabi H., Crivellari A., Homayouni S., Blaschke T., Ghamisi P., 2022a. Landslide detection using deep learning and object-based image analysis. Landslides, 19, 929–939. https://doi.org/10.1007/s10346-021-01843-x.

Ghorbanzadeh O., Xu Y., Zhao H., Wang J., Zhong Y., Zhao D., Zang Q., Wang S., Zhang F., Shi Y., Zhu X.X., Bai L., Li W., Peng W., Ghamisi P., 2022b. The Outcome of the 2022 Landslide4Sense Competition: Advanced Landslide Detection From Multisource Satellite Imagery. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens., 15, 9927–9942. https://doi.org/10.1109/JSTARS.2022.3220845.

Hang H.T., Hoa P.D., Tru V.N., Phuong N.V., 2021. Landslide Susceptibility Mapping Along National Highway-6, Hoa Binh Province, Vietnam Using Frequency Ratio Model And Gis. Int. J. Geomate. 21, 84–90. https://doi.org/10.21660/2021.85.j2222.

Hoang T.T.H., Nguyen Q.A., 2022. Assessing climate and hydrological conditions for tourism development in the Hoa Binh hydropower reservoir area, Hoa Binh province, in: Proceedings of the 13th National Geography Conference, p.11.

Hung P. Van, Son P.Q., Dung N. Van, 2015. The study evaluated arming of risk of lanslide in Hoa Binh and Son La reservoir hydropower area on the basis of analyzing high-resolution remote sensing and geographic information systems. Vietnam J. Earth Sci., 37(2), 193–203.

Khan R., Yousaf S., Haseeb A., Uddin M.I., 2021. Exploring a Design of Landslide Monitoring System. Complexity, 2021, 5552417, 13. https://doi.org/10.1155/2021/5552417.

Kubalíková L., Kirchner K., Bajer A., 2021. Geomorphological Resources for Geoeducation and Geotourism, in: Advances in Geographical and Environmental Sciences, 343–358. https://doi.org/10.1007/978-981-15-4956-4_18.

Lee Y.J., Jayakumar R., 2021. Economic impact of UNESCO Global Geoparks on local communities: Comparative analysis of three UNESCO Global Geoparks in Asia. Int. J. Geoheritage Park., 9, 189–198. https://doi.org/10.1016/j.ijgeop.2021.02.002.

Hoang, H.T.T., 2014. Multi-scale Analysis of Human-Environment Interactions. A Case-study in the Northern Vietnamese Mountains. Katholieke Universiteit Leuven, 12, 237.

Luu C., Ha H., Bui Q.D., Luong N.D., Khuc D.T., Vu H., Nguyen D.Q., 2023. Flash flood and landslide susceptibility analysis for a mountainous roadway in Vietnam using spatial modeling. Quat. Sci. Adv., 11, 100083. https://doi.org/10.1016/j.qsa.2023.100083.

Mohan A., Singh A.K., Kumar B., Dwivedi R., 2021. Review on remote sensing methods for landslide detection using machine and deep learning. Trans. Emerg. Telecommun. Technol., 32, 1–23. https://doi.org/10.1002/ett.3998.

Naveen D., Nirmala M., Roopesh D., Reddy J.K., Raju P.K., 2022. Landslide Detection Using Machine Learning Algorithms. J. Algebr. Stat., 13, 2822–2828.

Newsome D., Dowling R., 2018. Geoheritage and geotourism, Geoheritage: Assessment, Protection, and Management. Elsevier Inc., 305–321 https://doi.org/10.1016/B978-0-12-809531-7.00017-4.

Ngo V.L., Nguyen H., Dang K.B., Giang T.L., Dang V.B., Do T.H., Nguyen M.H., Dang N.V., Dao M.D., 2025. Advancing debris flow detection based on deep learning model and high-resolution images. Vietnam J. Earth Sci., 47(2), 290–214.

Nguyen B.D., Dang T.H., Vu D.M., Le T.T.H., 2011. Studying to determine causes of landslide in the area of the Mong Sen bridge, Lao Cai province. J. Earth Sci., 33(2), 164–174.

Nguyen H.D., Vu C.T., Bretcan P., Petrisor A.-I., 2024. Assessing the relationship between landslide susceptibility and land cover change using machine learning. Vietnam J. Earth Sci., 46(3), 339–359.

Nhu V.H., Hoang N.D., Nguyen H., Ngo P.T.T., Thanh Bui T., Hoa P.V., Samui P., Tien Bui D., 2020. Effectiveness assessment of Keras based deep learning with different robust optimization algorithms for shallow landslide susceptibility mapping at tropical area. Catena, 188, 104458. https://doi.org/10.1016/j.catena.2020.104458.

Pham B.T., Jaafari A., Nguyen D.D., Bayat M., Nguyen H.B.T., 2022. Development of multiclass alternating decision trees based models for landslide susceptibility mapping. Phys. Chem. Earth, 128, 103235. https://doi.org/10.1016/j.pce.2022.103235.

Pham V.T., Le H.L., Tran T.N., Nguyen Q.P., Phan T.T., Dinh T.Q., Dao M.D., Nguyen C.L., Nguyen H.C., 2023. Mechanism and numerical simulation of a rapid deep-seated landslide in Van Hoi reservoir, Vietnam. Vietnam J. Earth Sci., 45(3), 357–373.

Phong T.V., Ly H.-B., Trinh P.T., Prakash I., Hoan D.T., 2020. Landslide susceptibility mapping using Forest by Penalizing Attributes (FPA) algorithm based machine learning approach. Vietnam J. Earth Sci., 42(3), 237–246.

Prakash N., Manconi A., Loew S., 2021. A new strategy to map landslides with a generalized convolutional neural network. Sci. Rep., 11, 1–15. https://doi.org/10.1038/s41598-021-89015-8.

Roa-Lobo J.G., 2007. Identifying Landslide Hazards in a Tropical Mountain Environment, using Geomorphologic and Probabilistic Approaches. University of Maryland, 172.

Schweigl J., Hervas J., 2009. Landslide mapping in Austria, JRC Scientific and Technical Reports EUR 23785 EN. https://doi.org/10.2788/85150.

Sharma N., Saharia M., Ramana G. V., 2024. High resolution landslide susceptibility mapping using ensemble machine learning and geospatial big data. Catena, 235, 107653. https://doi.org/10.1016/j.catena.2023.107653.

Sun D., Ding Y., Wen H., Zhang F., Zhang Junyi, Gu Q., Zhang Jialan, 2024. SHAP-PDP hybrid interpretation of decision-making mechanism of machine learning-based landslide susceptibility mapping: A case study at Wushan District, China. Egypt. J. Remote Sens. Sp. Sci., 27, 508–523. https://doi.org/10.1016/j.ejrs.2024.06.005.

Tofani V., Segoni S., Agostini A., Catani F., Casagli N., 2013. Technical note: Use of remote sensing for landslide studies in Europe. Nat. Hazards Earth Syst. Sci., 13, 299–309. https://doi.org/10.5194/nhess-13-299-2013.

Tran A.T., Pham V.H., Tran T.T., Nguyen T.A.N., Nguyen V.D., Pham T.H., Tran V.P., 2024. Landslide susceptibility in Phuoc Son, Quang Nam: A deep learning approach. Vietnam J. Earth Sci., 47(1), 39–57.

Tricia R.S., Marian L., Burns W.J., Justin M., 2019. Preparing for Landslide Hazards a Land Use Guide for Oregon Communities, Oregon City GIS, 257.

Valchev N., Eftimova P., Andreeva N., Prodanov B., 2017. Application of Bayesian Network As a Tool for Coastal Flooding Impact Prediction At Varna Bay (Bulgaria, Western Black Sea). Coast. Eng. Proc., 1, 14. https://doi.org/10.9753/icce.v35.management.14.

World Bank Group, 2021. Economics for Disaster Prevention and Preparedness SUMMARY REPORT Investment in Disaster Risk Management in Europe Makes Economic Sense. World Bank, Washington, DC, USA.

World Bank Group, 2020. Seismic Resilience and Energy Efficiency in Public Buildings Project Information Document. Washington, DC.

Yang X., Chen L., 2010. Using multi-temporal remote sensor imagery to detect earthquake-triggered landslides. Int. J. Appl. Earth Obs. Geoinf., 12, 487–495. https://doi.org/10.1016/j.jag.2010.05.006.

Yao X., Yang H., Wu Y., Wu P., Wang B., Zhou X., Wang S., 2019. Land use classification of the deep convolutional neural network method reducing the loss of spatial features. Sensors (Switzerland), 19. https://doi.org/10.3390/s19122792.

Zhao C., Lu Z., 2018. Remote sensing of landslides-A review. Remote Sens., 10, 8–13. https://doi.org/10.3390/rs10020279.

Appendix A

```
Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  1.158e+01  8.772e-01  13.207  < 2e-16 ***
BSI          1.655e+01  1.333e+00  12.412  < 2e-16 ***
NDVI        -3.076e+00  2.687e-01 -11.450  < 2e-16 ***
NDBI        -6.162e+00  1.340e+00  -4.600 4.23e-06 ***
LULC         3.003e-01  3.278e-02   9.160  < 2e-16 ***
TIR          2.463e+00  4.027e-01   6.116 9.60e-10 ***
STI          6.073e-03  4.787e-03   1.269 0.204533
SPI          5.524e-02  3.117e-02   1.772 0.076392 .
Fault        2.256e+03  2.123e+02  10.629  < 2e-16 ***
TWI         -1.112e-01  4.107e-02  -2.709 0.006757 **
flowacc      2.783e-04  1.131e-04   2.460 0.013876 *
Pl_Cur       3.600e+04  2.405e+04   1.497 0.134453
Pr_Cur      -3.600e+04  2.405e+04  -1.497 0.134452
Curvature   -3.600e+04  2.405e+04  -1.497 0.134452
Slope        7.197e-02  5.021e-03  14.334  < 2e-16 ***
DEM          4.712e-04  2.149e-04   2.193 0.028316 *
Aspect       3.799e-04  4.138e-04   0.918 0.358633
River       -3.561e-03  7.158e-04  -4.974 6.55e-07 ***
Road        -2.389e-04  6.597e-05  -3.622 0.000292 ***
Pre_Avg     -1.137e-01  6.192e-03 -18.360  < 2e-16 ***
GeoType2     1.411e-01  1.208e-02  11.677  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

```
Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  1.217e+01  7.757e-01  15.691  < 2e-16 ***
BSI          1.678e+01  1.344e+00  12.488  < 2e-16 ***
NDVI        -2.900e+00  2.629e-01 -11.032  < 2e-16 ***
NDBI        -6.428e+00  1.356e+00  -4.740 2.14e-06 ***
LULC         2.926e-01  3.199e-02   9.146  < 2e-16 ***
TIR          2.776e+00  3.836e-01   7.237 4.58e-13 ***
SPI          8.274e-02  2.967e-02   2.789  0.00529 **
Fault        2.213e+03  2.101e+02  10.531  < 2e-16 ***
TWI         -7.684e-02  3.719e-02  -2.066  0.03883 *
flowacc      3.238e-04  1.121e-04   2.888  0.00388 **
Slope        7.356e-02  4.684e-03  15.705  < 2e-16 ***
River       -3.611e-03  7.028e-04  -5.138 2.78e-07 ***
Road        -2.108e-04  6.443e-05  -3.271  0.00107 **
Pre_Avg     -1.193e-01  5.699e-03 -20.931  < 2e-16 ***
GeoType2     1.408e-01  1.204e-02  11.689  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure 1.1. Elimination process to select suitable variables for landslide susceptibility assessment

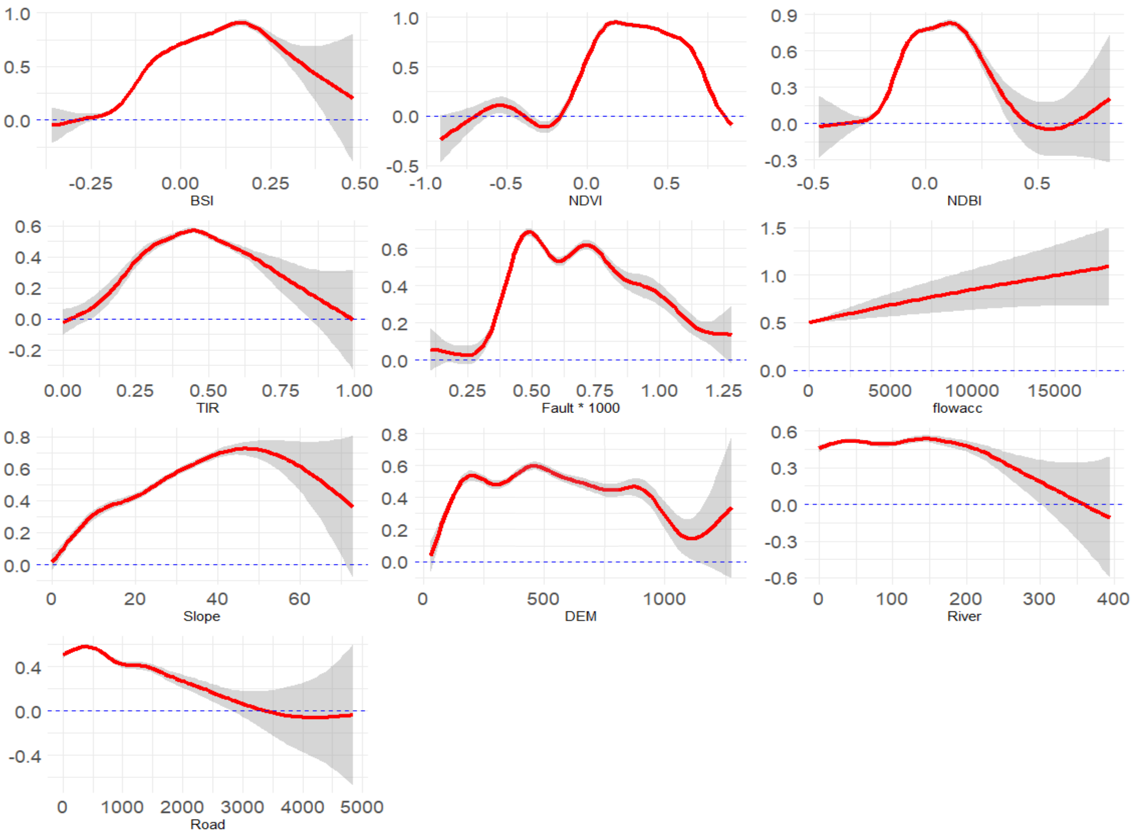


Figure 2.2. The relation between the most crucial conditioning factors to assess landslide susceptibility

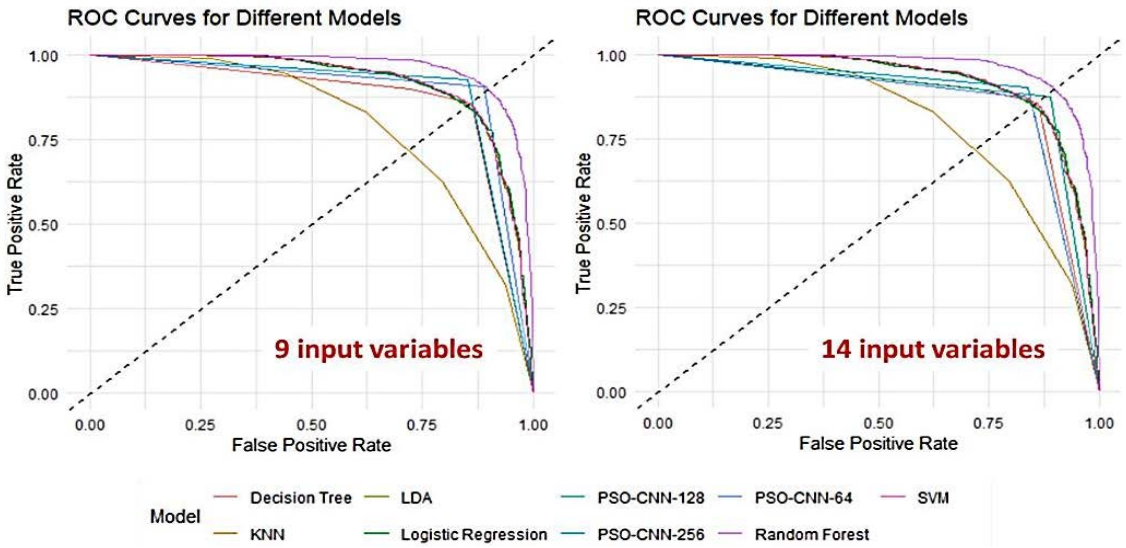


Figure 3.3. ROC curves of machine learning models using 9 and 14 variables in landslide susceptibility assessment

Appendix B: Brief description of deep learning model to detect landslide traces published in the paper titled: *"Deep learning models integrating multi-sensor and temporal remote sensing to monitor landslide traces in Vietnam"* of our research group, published in the "International Journal of Disaster Risk Reduction" journal in 2024.

The research contains three essential stages to follow. The first stage is the identification of landslide indicators. An in-depth analysis of field observations and satellite data was carried out as part of this process. Field investigators were used to identify tension cracks with topographical anomalies and vegetation patterns existing within landslide areas (Dang et al., 2024b). The authors used landslide indicators correctly identified in Worldview-2 satellite images to serve as effective training data for DL models.

Training data preparation constitutes the second stage of the procedure. Research investigators specifically chose six Vietnamese regions prone to landslides (Yang and Chen, 2010). A group of landslide-prone areas in Vietnam was documented through high-resolution satellite imagery obtained using Google Earth Pro's Worldview-2 technology. The authors divided the collected images into smaller ones to increase the practical usage of the training and validating data. The research method served to identify landslide indicators on the satellite imagery.

Based on an optimal input database, the authors achieved success in designing a robust DL model in the third stage by implementing a complete data preparation strategy. DeepLab-v3 with ResNet-34 setup demonstrated the highest effectiveness in detection tasks through its 0.213 Dice coefficient and 96.8% accuracy (Dang et al., 2024a; Ngo et al., 2025). The authors prevented overfitting by implementing a testing dataset, even though they trained with validation data. The input size of 256×256 images enables better data processing for classification, but it also causes simplified landslide traces to become indistinguishable. The detection of large-scale motion was more accurate for 64×64 images, although it failed to capture details (Dang et al., 2022). DeepLab-v3 with ResNet backbones displayed superior performance to the combination of VGG11 and PSPNet. In landslide detection applications, DeepLab-v3 proves to be an essential DL approach that successfully identifies traces according to Yao et al (2019). Once the training model was completed, a continuous operation mode enabled DeepLab-v3 to identify landslides that had not been involved in its training phase (Chen et al., 2018).

Appendix C: Description of each type of landslide conditioning factors

Firstly, lithological and geological factors, including characteristics of weathering crust and fault density, play a significant role in determining the likelihood of a landslide (Borrelli et al., 2014). Strong, solid materials are less likely to cause landslides compared to weak, fragmented ones. The proximity to faults and lineaments, which are weak points in the Earth's crust, also increases landslide hazard (Tu et al., 2016). Areas close to these zones are more prone to instability and breaking. In a plan to build a new road through a mountainous region, a road-construction company conducted a geological survey and found that one proposed route passes through an area with numerous lineaments (Ghasemian et al., 2020). The scientists might modify the roads when they encounter weak geological formations, ensuring safety for the road while also protecting residents and visitors.

The properties of geologic strata serve as fundamental indicators to determine the chances of slope failure developing, according to Malet and Maquaire (2012) and Ren (2015). Fine-grained clay soils retain more water than coarse-grained sand soils because clay holds water more effectively, yet sand drains better. Therefore, the clay-based soils have higher landslide risk potential. By including information about soil roughness, the research can yield additional findings, as this factor determines how water interacts with soil and affects the displacement process (Tarolli et al., 2014). The comprehensive study of these factors improves predictive accuracy in models that lead to better landslide prevention methods.

Secondly, geomorphological factors, such as slopes with high angles, tend to collapse due to the enhanced gravitational force acting upon them (V. A. Tran et al., 2024). Slopes with reversed curved (concave) forms attract water and debris, increasing their instability levels, whereas slopes with forward curved (convex) forms demonstrate better stability. Therefore, the Sediment Transport Index (STI) uses topographical features to measure sediment mobility by evaluating surface area along with slope gradient. Higher STI values suggest a greater likelihood of landslides due to material movement. The Terrain Ruggedness Index (TRI) indicates land complexity, with rougher terrain generally having a higher risk of landslides.

Thirdly, hydrological factors, such as the Topographic Wetness Index (TWI), are crucial in determining the amount of water in the soil (Pham et al., 2022). Firstly, TWI provides slope water accumulation evaluation and produces elevated indicators that indicate increased landslide risk. The Stream Power Index (SPI) evaluates stream properties to detect erosion potential through its values, which rise as landslide threats expand in streamside areas. The analysis requires consideration of two critical factors, which include the distance to streams as well as the rainfall intensity. The risk for erosion is elevated in areas adjacent to streams. The effects of heavy rainfall include soil saturation, which pushes up water pressure and produces landslides throughout the territory. Weather records serve to improve the analysis when they are made available.

Fourthly, precipitation acts as a major triggering factor in landslides, which can cause sudden instability of slopes in a very short time. Quantifying and analyzing the relationship between rainfall and landslide occurrence is a crucial foundation for developing effective early warning models. This research utilized data collected by the National Hydrometeorological Service's monitoring system, which included records spanning multiple years, both daily and monthly. Rainfall was grouped into (i) the periods when landslides occur (generally June to September) and (ii) those with no landslides. At the actual landslide events - collected from remote sensing and field surveys - rainfall information is replicated and specifically labeled as "landslide-triggering rainfall". This replication enhances the training sample with input signals characteristic of high-risk rainfall events, while providing a clear context for the DL model to differentiate between normal and dangerous rainfall. This approach not only helps improve the sensitivity and accuracy of the model but also aims at the long-term goal of building a real-time landslide prediction system where rainfall is integrated as a key input signal, contributing to improving early warning capabilities and reducing disaster risks in vulnerable mountainous areas.

Lastly, anthropogenic influences, which refer to human activities, have a substantial impact on the stability of slopes (Kayastha et al., 2013). Changes in land use may modify the likelihood of landslides. Plant root systems, together with other vegetation, scale down slope deterioration and fortify earth materials, so that landslides occur less frequently. Landslide frequency increases when vegetation disappears due to activities such as deforestation or specific farming systems, as soil stability decreases. Kinds of human activities that either raise slopes' weight limits or cause water flow changes make landslides more dangerous. Highway development (e.g., opening new roads, building homestays) serves as a key demonstration for tourism development because it has the power to modify regional water drainage systems and might destabilize adjacent slopes. The positions nearest to roads experience higher landslide vulnerability due to their location.

Appendix D: Details about machine learning development for mapping landslide susceptibility

1. Random Forest (RF)

Random Forest demonstrates excellent capability when processing data with high dimensions, often applied in landslide susceptibility mapping (Abdo and Richi, 2024). As an ensemble method, Random Forest operates by lowering information variance and produces models that survive data errors effectively. The algorithm can forecast landslide occurrences with unlisted information through these trends. Random Forest constructs an ensemble model through the combination of many decision trees (Le et al., 2022). This ensemble approach leverages the predictive power of multiple trees. The development of the forest originates from picking random predictor variables (features) throughout the training data. The introduction of randomness reduces model variance while preventing it from fitting exclusively to the training data, which in turn causes the model to lose its ability to generalize (Sannigrahi et al., 2019). The ensemble prediction of Random Forest trees improves data quality by averaging multiple trees, which reduces overfitting and eliminates data anomalies. The model achieves better stability and a wider application range through this averaging technique. Random Forest shows exceptional capability to determine intricate patterns between different variables that impact landslide susceptibility. The algorithm uses many predictor variables alongside established landslide locations to find relationships between various variables.

The authors performed parameter fine-tuning of the Random Forest model for constructing landslide susceptibility maps. The crucial elements that guide forest prediction include both tree depth and the number of trees used. The model stability and variance reduction benefit from more tree implementation in the system at the cost of lengthened training times and increased computational expenses (Pham et al., 2024). Research-based assessments of different hyperparameters determine the best selection of optimal trees. Caret operates through "Maximum Depth" to determine maximum tree depth levels. Deep decision trees detect advanced relationships in data, but they become more likely to create an overfitting model. Advantageously shallow trees avoid overfitting but can fail to identify distinct patterns. The maximum depth value needs careful adjustment because it supports model performance retention. Therefore, the authors selected 1000 branches as the most suitable value after conducting their tests. The automatic selection of the best model occurred during the loop test process.

2. Support Vector Machines (SVM)

SVM models represent a strong ML technique that helps generate landslide susceptibility maps by Bui et al. (2012b). The SVM models establish the optimal hyperplane, which identifies the most separate position between landslide data points and non-landslide groups with maximum space between them. The margin in support vector machines contains support vectors as its boundary points that represent data points closest to the hyperplane in each category (Bui et al., 2012). SVMs achieve accurate prediction results on unobserved data because they can determine the optimal hyperplane that separates the data. The mapping process for landslide influencing factors utilizes higher-dimensional spaces while kernel functions reveal hidden nonlinear attribute relations between features. The methodology maintains the broad applicability of the model through space optimization, thereby avoiding specialization for the training data. Feature scaling or normalization techniques should be used with SVMs because prominent features often dominate model decisions. Random Forest shares a similar advantage with class scaling, as it enhances model efficiency when modeling unbalanced data samples.

During SVM training, three essential parameters must be optimized, starting with the Gamma Parameter (γ) and moving to the Cost Parameter (C) and ending with the Kernel Function (Nguyen et al., 2022). The kernel function defines both the data transformation process and the SVM separation methods. Radial basis functions, linear functions, and polynomial functions represent the standard kernel functions in SVM implementation. Groups choose their kernel function after thorough data relationship analysis, which results in reflective experimental determination. The Cost Parameter (C) helps achieve the best performance balance between error prevention and creating a large margin. The margin increases with greater values of C, although it can potentially trigger overfitting conditions. When the C value decreases, classification errors become more probable; yet, the model demonstrates better generalizing capabilities. The Gamma Parameter (γ) enables a controller mechanism to determine which data points will shape the decision boundary. When gamma increases in magnitude, the nearby data points become more influential, thus causing the decision boundary to become more intricate. Smoothing of the decision boundary occurs when the gamma parameter is set low since it lessens the effects of single points on its structure. The authors selected 90 as the C value plus γ at 0.4 as the ideal parameters for this analysis. SVMs achieve effective landslide susceptibility mapping when these adjustable parameters are adjusted to predict landslide hazards based on various influencing factors.

3. K-Nearest Neighbors (KNN)

The KNN model operates as a user-friendly and effective system for creating landslide susceptibility maps (Nguyen et al., 2024). The system examines newly-introduced data points by evaluating their associations to specific labels present among the k closest neighboring training samples. The KNN model preserves all attributes of training data points accompanied by their designated landslide classification.

Based on new data entries, the model determines its distances from its k closest neighbors in the training set through measurements such as Euclidean distance (Ma et al., 2021). The model classifies new data points as either landslide or non-landslide through the decision of its k nearest neighbor members. The classification decision for new data points emerges from the combined influence of training data points that most closely resemble it. The model's performance can be improved when dealing with imbalanced landslide data by combining class scaling with less sensitive distance measures for the main class. The successful training of KNN requires a careful selection between the number of neighbors (k) and the appropriate distance measure. During classification operations, k defines the number of near neighbors that the system will assess. A high k value minimizes data noise yet causes less accurate classifications across the dataset. Data patterns that occur locally are more likely to be identified when K values remain small, yet risks exist because of noise contamination. Cross-validation is used in this research to determine the suitable k value selection.

4. Linear Discriminant Analysis (LDA)

The traditional usage of LDA models enables researchers to map landslide susceptibility (Pham et al., 2016). The separation criteria for linear data triggers LDA to conduct effective class division of two or more distinct data points. The processing requirements of LDA make it suitable for handling extensive datasets, as it requires minimal computational power the procedure known as LDA functions to reduce datasets that contain multiple features. Through identifying the fundamental variations between different classes, LDA produces a subset of essential features that enhances classification effectiveness. In landslide susceptibility mapping, areas are categorized into landslide-prone and non-landslide-prone regions, which are then compared based on various input factors influencing them. In this study, the objective is to identify linear changes that minimize differences within "landslide" and "non-landslide" classes while maximizing separation between them. A dimensionality reduction technique called Principal Component Analysis (PCA), combined with other approaches, should be used to preprocess datasets with many features before executing LDA. PCA identifies uncorrelated feature subsets that describe most of the original data variations to enhance the overall model performance. The evaluation of input data must be done thoroughly to maintain proper adherence to regularity and separability principles for landslide hazards.

5. PSO-CNN

Syulistyo et al. (2016) developed PSO-CNN by integrating the Particle Swarm Optimization (PSO) algorithm with Convolutional Neural Network (CNN) for their application. This particular method combines advantages from both techniques to boost CNN performance and operational efficiency, specifically when used in fields involving image processing and computer vision. The PSO is a stochastic optimization approach that is based on the social behavior of birds flocking or fish schooling. In the PSO, a group of particles iteratively explores a multidimensional search space to optimize a specific objective function. Each particle effectively searches and utilizes the search space by using its own most optimal known location and the best-known position of neighboring particles. Whereas CNNs are a specific sort of deep neural network designed specifically for processing structured matrices. CNNs construct layers of interconnected neurons by using convolutional operations, pooling, and nonlinear activations. This enables them to learn hierarchical representations of the input data. Such a design lets them create a hierarchical representation of input data, which results in learning. The network generates substantial performance outcomes when utilized in image segmentation, along with object detection and image classification operations.

To assess the landslide susceptibility in this study, the PSO-CNN was chosen to potentially optimize crucial parameters, including weights, biases, and architectural choices (Devarakonda and Bozic, 2016). The typical training of CNNs primarily relies on gradient-based algorithms, particularly with stochastic gradient descent (SGD). However, SGD got trapped in local minima or encountered difficulties in determining the optimal hyperparameters. PSO-CNN addresses these challenges by using the global search capability of PSO to discover enhanced combinations of CNN parameters. The designed CNN utilizes ReLU activation functions in combination with three 1D convolutional layers, which must precede max pooling layers. The model proceeds with a flatten layer that reshapes its 2D output to create a one-dimensional vector. Two dense layers follow the previous ones before the SoftMax activation function finishes the classification process. By adjusting its hyperparameters, namely the learning rate and batch size, the performance of the CNN may be improved (Wang et al., 2020). The technique handles CNN model construction followed by its training process using specified parameters. The algorithm obtains the negative maximum validation accuracy before using it as an optimization target for Particle Swarm Optimization (PSO). The PSO optimization strategy searches the hyperparameter space to identify the optimal learning rate values, along with the best batch size values. The CNN model was trained in 100 loops using the identified optimum hyperparameters, and its performance was evaluated on validation data once these parameters were determined. The model's predictions are converted into class labels and assessed using metrics such as the area under the Curve (AUC) of the ROC curve and a confusion matrix. These values provide a thorough evaluation of the model's classification accuracy and are explained in the next section.

Appendix E: The description of each measurement point includes the following definition

The beginning step determines accuracy by dividing correct predictions by the total number of evaluation data samples. The authors utilized this method under expected conditions where data classes remain well-balanced amongst equal sample numbers. The accuracy value emerges from this calculation formula:

$$ACC = \frac{Number\ of\ correct\ predictions}{Total\ number\ of\ samples} \quad (1)$$

Model precision enables better prediction of actual labels from expected outcomes, thus providing a general evaluation of model effectiveness. Additionally, the statistical metric known as Kappa determines the actual measurement accuracy in reference to potential random precision levels. The computation is as follows:

$$K = \frac{P_{observed} - P_{chance}}{1 - P_{chance}} \quad (2)$$

The Kappa equation demonstrates the relationship between $P_{observed}$ and P_{chance}

shows the actual accuracy and the expected accuracy from random selection. A Kappa score approaching 1 aligns with high accuracy in the model, but a score approaching 0 establishes that agreement aligns with random chance levels. The calculation helps evaluate the dependability of the model beyond its numerical accuracy levels.

The model demonstrates enhanced landslide detection capabilities, as it delivers superior sensitivity values (equivalent to the true positive rate), indicating its capacity to recognize actual events accurately. The sensitivity calculation requires the application of this formula:

$$\text{Sensitivity} = \frac{True\ positives}{True\ positives - False\ negatives} \quad (3)$$

Specificity measures the True Negative Rate, which indicates how well the model recognizes genuine negative instances. This research demonstrates the model's capacity to accurately detect areas that do not have a landslide risk. The calculation depends on this mathematical formula:

$$\text{Specificity} = \frac{True\ negatives}{True\ negatives - False\ positives} \quad (4)$$

The conditional accuracy of positive predictive value (PPV) determines how many correct forecasts exist among all optimistic predictions from the model. The model demonstrates reliable accuracy in landslide prediction when the PPV value is high, as this indicates accurate optimistic predictions. The calculation for PPV contains the following mathematical relationship:

$$\text{PPV} = \frac{True\ positives}{True\ negatives + False\ positives} \quad (5)$$

When calculating Negative Predictive Value, the percentage of negative predictions that turn out to be accurate is measured. The model demonstrates high reliability in negative prediction because it accurately identifies areas free from landslides according to a larger value. The calculation of NPV value uses this equation:

$$\text{NPV} = \frac{True\ negatives}{True\ negatives + False\ positives} \quad (6)$$

Lastly, the AUC-ROC measurement defines the area under the curve of the Receiver Operating Characteristic (ROC). The ROC curve plots the true positive rate against the false positive rate at different threshold values. Models display strong abilities in predictions between positive and negative samples when their AUC values change from zero to one.